

GenCore version 5.1.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 21, 2005, 19:59:00 ; Search time 241 Seconds
(without alignments)

2728.433 Million cell updates/sec

Title: US-10-079-429A-4

Perfect score: 4812

Sequence: 1 MKQLPAATVRLSSQIIITS.....KECVHGRPPFHLTYLPETT 932

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters: 2166443

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

UniProt_05.80.*

1: uniprot_sprot.*

2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4812	100.0	932	1 PMS1_HUMAN	P54277 homo sapien
2	4742	98.5	920	2 Q4VAL4_HUMAN	Q4VAL4 homo sapien
3	4693.5	97.5	931	2 Q5R904_PONPY	Q5R904 pongo pygma
4	4584.5	95.3	893	2 Q5FBZ3_HUMAN	Q5FBZ3 homo sapien
5	3889	80.8	770	2 Q5FBZ8_HUMAN	Q5FBZ8 homo sapien
6	3579.5	74.4	917	2 Q8K119_MOUSE	Q8K119 mus musculus
7	3543.5	73.6	919	2 Q6P7D0_RAT	Q6P7D0 rattus norv
8	3351	69.6	669	2 Q4VAL5_HUMAN	Q4VAL5 homo sapien
9	3349	69.6	667	2 Q5FBZ9_HUMAN	Q5FBZ9 homo sapien
10	2779.5	57.8	555	2 Q5FBZ6_HUMAN	Q5FBZ6 homo sapien
11	2720	56.5	916	2 Q5ZKT5_CHICK	Q5ZKT5 gallus gall
12	2319	48.2	925	2 Q7ZXV9_XENLA	Q7ZXV9 xenopus lae
13	2213.5	46.0	928	2 Q5FVX9_XENLA	Q5FVX9 xenopus tro
14	2046.5	42.5	896	2 Q8JFR9_BRARE	Q8JFR9 brachydanio
15	1835.5	38.1	854	2 Q4RTJ3_TETNG	Q4RTJ3 tetraodon n
16	1177	24.5	234	2 Q5FBZ2_HUMAN	Q5FBZ2 homo sapien
17	1132.5	23.5	372	2 Q7SKD5_BRARE	Q7SKD5 brachydanio
18	977	20.3	248	2 Q5FBZ4_HUMAN	Q5FBZ4 homo sapien
19	976	20.3	196	2 Q5FBZ5_HUMAN	Q5FBZ5 homo sapien
20	975	20.3	195	2 Q5FBZ1_HUMAN	Q5FBZ1 homo sapien
21	830	17.2	163	2 Q6BDF0_HUMAN	Q6BDF0 homo sapien
22	719	14.9	194	2 Q8JFW5_BRARE	Q8JFW5 brachydanio
23	708	14.7	166	2 Q96H10_HUMAN	Q96H10 homo sapien
24	702	14.6	165	2 Q5XG96_HUMAN	Q5XG96 homo sapien
25	659.5	13.7	880	2 Q7Q1Y1_ANOGA	Q7Q1Y1 anopheles g
26	654	13.6	1022	2 Q54QA0_DICDI	Q54QA0 dictyosteli
27	639	13.3	143	2 Q8BLI9_MOUSE	Q8BLI9 mus musculus
28	616.5	12.8	895	2 Q8T9C0_DROME	Q8T9C0 drosophila
29	615.5	12.8	899	2 Q9V7B6_DROME	Q9V7B6 drosophila
30	607	12.6	923	2 Q941I6_ARATH	Q941I6 arabidopsis
31	595.5	12.4	893	2 Q76417_DROME	Q76417 drosophila

32 587.5 12.2 805 2 Q9TVL8_CABEL Q9TVL8 caenorhabdi
33 584 12.1 871 2 Q5ZJ94_CHICK Q5ZJ94 gallus gall
34 581.5 12.1 903 2 / Q755U7_ASHGO Q755U7 ashbya goss
35 572.5 11.9 797 2 Q60M36_CABER Q60M36 caenorhabdi
36 569 11.8 862 1 PMS2_HUMAN P54278 homo sapien
37 569 11.8 862 2 Q75MR2_HUMAN Q75MR2 homo sapien
38 569 11.8 879 2 Q8NSQ6_HUMAN Q8NSQ6 homo sapien
39 568 11.8 862 2 Q52LH6_HUMAN Q52LH6 homo sapien
40 564.5 11.7 907 2 Q6PPA0_CANGA Q6PPA0 candida gla
41 563 11.7 923 2 Q69L72_ORYSA Q69L72 oryza sativ
42 561.5 11.7 893 2 Q6C6B8_YARLI Q6C6B8 yarrowia li
43 561 11.7 908 2 -Q8TG50_YEAST Q8TG50 saccharomyc
44 559.5 11.6 1094 2 Q4XWC3_PLACH Q4XWC3 plasmodium
45 558 11.6 908 2 Q8TG48_YEAST Q8TG48 saccharomyc

ALIGNMENTS

RESULT 1
ID PMS1_HUMAN STANDARD; PRT; 932 AA.
AC P54277;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 10-MAY-2005 (Rel. 47, Last annotation update)
DE PMS1 protein homolog 1 (DNA mismatch repair protein PMS1).
GN Name=PMS1; Synonyms=PMS1L;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC Homo.
OX NCBI_TaxID=9606;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Gall bladder;
RX MEDLINE=94352394; PubMed=8072530; DOI=10.1038/371075a0;
RA Nicolaides N.C., Papadopoulos N., Liu B.-Y., Wei Y.-F., Carter K.C.,
RA Ruben S.M., Rosen C.A., Haseltine W.H., Fleischmann R.D., Fraser C.M.,
RA Adams M.D., Venter J.C., Dunlop M.G., Hamilton S.R., Petersen G.M.,
RA de la Chapelle A., Vogelstein B., Kinzler K.W.;
RT "Mutations of two PMS homologues in hereditary nonpolyposis colon
cancer.";
RT Nature 371:75-80 (1994).
RN [2]
RP NUCLEOTIDE SEQUENCE [GENOMIC DNA], AND VARIANTS GLN-27; LYS-202;
RP ARG-501; SER-632; ASP-720 AND HIS-793.
RA Rieder M.J., Livingston R.J., Daniels M.R., Chung M.-W.,
RA Miyamoto K.E., Nguyen C.P., Nguyen D.A., Poel C.L., Robertson P.D.,
RA Schackwitz W.S., Sherwood J.K., Witrak L.A., Nickerson D.A.;
RT "NIHES-SNP, environmental genome project, NIHES ES15478, Department
of Genome Sciences, Seattle, WA (URL: http://egp.gs.washington.edu).";
RL Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.
RN [3]
RP VARIANTS HNPCC3 THR-394 AND ARG-501.
RX MEDLINE=99408236; PubMed=10480359; DOI=10.1007/s004390051067;
RA Wang Q., Lassus C., Desseigne P., Saurin J.-C., Maugard C., J.-F.,
RA Navarro C., Ruano E., Descos L., Trillet-Lenoir V., Boeset J.-F.,
RA Puisieux A.;
RT "prevalence of germline mutations of hMLH1, hMSH2, hPMS1, hPMS2, and
hMSH6 genes in 75 French kindreds with nonpolyposis colorectal
cancer";
RL Hum. Genet. 105:79-85 (1999).
CC -!- FUNCTION: Probably involved in the repair of mismatches in DNA.
CC -!- SUBCELLULAR LOCATION: Nuclear (Potential).
CC -!- DISEASE: Defects in PMS1 are the cause of hereditary non-polyposis
CC colorectal cancer type 3 (HNPCC3) [MIM:600258]. Mutations in more
CC than one gene locus can be involved alone or in combination in the
CC production of the HNPCC phenotype (also called Lynch syndrome).
CC Most families with clinically recognized HNPCC have mutations in
CC either MLH1 or MSH2 genes. HNPCC is an autosomal, dominantly
CC inherited disease associated with marked increase in cancer
CC susceptibility. It is characterized by a familial predisposition

to early onset colorectal carcinoma (CRC) and extra-colonic cancers of the gastrointestinal, urological and female reproductive tracts. HNPCC is reported to be the most common form of inherited colorectal cancer in the Western world, and accounts for 15% of all colon cancers. Cancers in HNPCC originate within benign neoplastic polyps termed adenomas. Clinically, HNPCC is often divided into two subgroups. Type I: hereditary predisposition to colorectal cancer, a young age of onset, and carcinoma observed in the proximal colon. Type II: patients have an increased risk for cancers in certain tissues such as the uterus, ovary, breast, stomach, small intestine, skin, and larynx in addition to the colon. Diagnosis of classical HNPCC is based on the Amsterdam criteria: 3 or more relatives affected by colorectal cancer, one a first degree relative of the other two; 2 or more generation affected; 1 or more colorectal cancers presenting before 50 years of age; exclusion of hereditary polyposis syndromes. The term "suspected HNPCC" or "incomplete HNPCC" can be used to describe families who do not or only partially fulfill the Amsterdam criteria, but in whom a genetic basis for colon cancer is strongly suspected.

--- SIMILARITY: Belongs to the DNA mismatch repair mutL/hexB family. ---
 --- SIMILARITY: Contains 1 HMG box DNA-binding domain. ---
 --- DATABASE: NAME=Hereditary non-polyposis colorectal cancer db; WWW="http://www.nfcdt.nl/" ---

--- This Swiss-Prot entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use as long as its content is in no way modified and this statement is not removed. ---

EMBL; UL3695; AAA63922.1; -; Genomic DNA.
 DR EMBL; AY267352; AAO89079.1; -; Genomic DNA.
 DR PIR; S47597; S47597.
 DR HSSP; PS4278; LH7S.
 DR ENSG0000064933; Homo sapiens.
 DR HGNC; HGNC:9121; PMS1.
 DR MIM; 600258; -.
 DR GO; GO:0005634; C:nucleus; TAS.
 DR GO; GO:0003677; F:DNA binding; TAS.
 DR GO; GO:006298; P:mismatch repair; TAS.
 DR InterPro; IPR003594; ATP bd ATPase.
 DR InterPro; IPR002099; DNA mis repair.
 DR InterPro; IPR000910; HMG 12 box.
 DR PANTHER; PTHR10073; DNA mis repair; 1.
 DR Pfam; PF01119; DNA_mis_repair; 1.
 DR Pfam; PF02518; HMG box; 1.
 DR TIGRPFAMS; TIGR00585; mutL; 1.
 DR PROSITE; PS00058; DNA mismatch repair_1; 1.
 DR PROSITE; PS50118; HMG_BOX_2; 1.
 DR Anti-oncogene; Cell cycle; Disease mutation; DNA damage; DNA repair; Hereditary nonpolyposis colorectal cancer; Nuclear protein; Polymorphism.
 KW DNA BIND HMG box.
 KW VARIANT 27 639 E -> Q (in dbSNP:5742973).
 FT 27 /FTid=VAR_019166.
 FT 202 R -> K (in dbSNP:2066459).
 FT /FTid=VAR_014877.
 FT 394 M -> T (in incomplete HNPCC3; dbSNP:1145231).
 FT /FTid=VAR_012967.
 FT 501 G -> R (in incomplete HNPCC3; dbSNP:1145232).
 FT /FTid=VAR_012968.
 FT 632 N -> S (in dbSNP:2066456).
 FT /FTid=VAR_014878.
 FT 720 E -> D (in dbSNP:2066455).
 FT /FTid=VAR_014879.
 FT 793 Y -> H (in dbSNP:1145234).
 FT /FTid=VAR_014880.
 FT SEQUENCE 932 AA; 105830 MW; EC4F402937B616DF CRC64;

Query Match 100.0%; Score 4812; DB 1; Length 932;
 Best Local Similarity 100.0%; Pred. No. 1.2e-214;
 Matches 932; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKQLPAATVRLSSQIITSVVVKELIENSLDAGATSDVKLENYGDKLEVRDNGSG 60
 DB 1 MKQLPAATVRLSSQIITSVVVKELIENSLDAGATSDVKLENYGDKLEVRDNGSG 60

QY 61 IKAVDAPVMAMKYTTSKINSHEDLENLTGYPRGALGSIICIAEVLITTRTAADNFSTQ 120
 DB 61 IKAVDAPVMAMKYTTSKINSHEDLENLTGYPRGALGSIICIAEVLITTRTAADNFSTQ 120

QY 121 VYLDGSHLSQKPSHLGGTGTVALRLFKNLVPRKQFYSTAKCKCKDEIKKIQDLMSFG 180
 DB 121 VYLDGSHLSQKPSHLGGTGTVALRLFKNLVPRKQFYSTAKCKCKDEIKKIQDLMSFG 180

QY 181 ILKPLRIIVFVNKAVIQKSRVSDHKALMSVLGTAVNNMESQYHSEESQIYLSGFL 240
 DB 181 ILKPLRIIVFVNKAVIQKSRVSDHKALMSVLGTAVNNMESQYHSEESQIYLSGFL 240

QY 241 PKCDADHSFTSLSTPERSFIFINSRPVHQKDLILIRHHYNLCKLKSTRLYPVFELKID 300
 DB 241 PKCDADHSFTSLSTPERSFIFINSRPVHQKDLILIRHHYNLCKLKSTRLYPVFELKID 300

QY 301 VFTADVVDNLTPDKSQVLLQNKESVLIAlENLMTTCYGPSPSTSYNNKTDVSAADIVL 360
 DB 301 VFTADVVDNLTPDKSQVLLQNKESVLIAlENLMTTCYGPSPSTSYNNKTDVSAADIVL 360

QY 361 SKTATDVLFNKVGSGKYNVDTSVTPFQNDMDNDSGKNTDCLNHOISIGDRGYGH 420
 DB 361 SKTATDVLFNKVGSGKYNVDTSVTPFQNDMDNDSGKNTDCLNHOISIGDRGYGH 420

QY 421 CSSEISNIDKNTKNAFQDISMSNVSWNSQTSYKTCFPISSVKHTQSENGKDHIDESGE 480
 DB 421 CSSEISNIDKNTKNAFQDISMSNVSWNSQTSYKTCFPISSVKHTQSENGKDHIDESGE 480

QY 481 NEEZEAGLENSSEISADEWSRGNILKNSVGENIEPVKILVPEKSLPCKVSNNNYPIPEQWN 540
 DB 481 NEEZEAGLENSSEISADEWSRGNILKNSVGENIEPVKILVPEKSLPCKVSNNNYPIPEQWN 540

QY 541 LNEDSCNKSNVIDNKGKVTAYDILLNRIKPKMSASALFVQDHRPQZLIENPKTSLED 600
 DB 541 LNEDSCNKSNVIDNKGKVTAYDILLNRIKPKMSASALFVQDHRPQZLIENPKTSLED 600

QY 601 ATLQIEELWKLTLSEBEKLYEKA TKDLERYNSQMKRAIQESQMSLKDGRKKIKPTSAW 660
 DB 601 ATLQIEELWKLTLSEBEKLYEKA TKDLERYNSQMKRAIQESQMSLKDGRKKIKPTSAW 660

QY 661 NLAQKHKLKTSLSNOPKLDLQSQIEKRSQNIKMVQIPFSMKNLKNFKKQNKVDLEE 720
 DB 661 NLAQKHKLKTSLSNOPKLDLQSQIEKRSQNIKMVQIPFSMKNLKNFKKQNKVDLEE 720

QY 721 KDEPCLINLPRPDPAWLTSTKTEVMLNPNRYVEEALLFKRLLENHKLPAEPLEKPTMLTE 780
 DB 721 KDEPCLINLPRPDPAWLTSTKTEVMLNPNRYVEEALLFKRLLENHKLPAEPLEKPTMLTE 780

QY 781 SLFNGSHYLDVLYKMTADDQRYSGSTYLSDPRLTANGFKILIPGVSIYENYLEIGMAN 840
 DB 781 SLFNGSHYLDVLYKMTADDQRYSGSTYLSDPRLTANGFKILIPGVSIYENYLEIGMAN 840

QY 841 CLPFTGVADLKEILNALINRNAKEVYECPRKVISYLEGEAVRLSRQLPWLKEDIQDI 900
 DB 841 CLPFTGVADLKEILNALINRNAKEVYECPRKVISYLEGEAVRLSRQLPWLKEDIQDI 900

QY 901 IYRMKHQFGNEIKECVCHGRPPFFHLLTYLPETT 932
 DB 901 IYRMKHQFGNEIKECVCHGRPPFFHLLTYLPETT 932

RESULT 2
 Q4VAL4 HUMAN
 ID Q4VAL4 HUMAN PRELIMINARY; PRT; 920 AA.
 AC Q4VAL4;

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 21, 2005, 19:59:00 ; Search time 241 Seconds
(without alignments)
2728.433 Million cell updates/sec

Title: US-10-079-429A-4
Perfect score: 4812
Sequence: 1 MKQLPAATVRLSSQLITS.....KECVHGRPFPHLTYPETT 932

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters: 2166443

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : UniProt 05.80.*

1: uniprot_sprot.*

2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4812	100.0	932	1 PMS1_HUMAN	P54277 homo sapien
2	4742	98.5	920	4 QVAL14_HUMAN	Q4val14 homo sapien
3	4693.5	97.5	931	2 QSR904_PONPY	Q5r904 pongo pygma
4	4584.5	95.3	893	2 Q5FBZ3_HUMAN	Q5fbz3 homo sapien
5	3889	80.8	770	2 Q5FBZ8_HUMAN	Q5fbz8 homo sapien
6	3579.5	74.4	917	2 Q8K119_MOUSE	Q8k119 mus musculus
7	3543.5	73.6	919	2 Q6P7D0_RAT	Q6p7d0 rattus norv
8	3351	69.6	669	2 Q4VAL5_HUMAN	Q4val5 homo sapien
9	3349	69.6	667	2 Q5FBZ9_HUMAN	Q5fbz9 homo sapien
10	2779.5	57.8	555	2 Q5FBZ6_HUMAN	Q5fbz6 homo sapien
11	2720	56.5	916	2 Q5ZKT5_CHICK	Q5zkt5 gallus gall
12	2319	48.2	925	2 Q7ZXV9_XENLA	Q7zxv9 xenopus lae
13	2213.5	46.0	928	2 Q5FVX9_XENTR	Q5fvx9 xenopus tro
14	2046.5	42.5	896	2 Q8JFR9_BRARE	Q8jfr9 brachydanio
15	1835.5	38.1	954	2 Q4RTJ3_TETNG	Q4rtj3 tetraodon n
16	1177	24.5	234	2 Q5FBZ2_HUMAN	Q5fbz2 homo sapien
17	1132.5	23.5	372	2 Q7SXD5_BRARE	Q7sxd5 brachydanio
18	977	20.3	248	2 Q5FBZ4_HUMAN	Q5fbz4 homo sapien
19	976	20.3	196	2 Q5FBZ5_HUMAN	Q5fbz5 homo sapien
20	975	20.3	195	2 Q5FBZ1_HUMAN	Q5fbz1 homo sapien
21	830	17.2	163	2 Q68DF0_HUMAN	Q68df0 homo sapien
22	719	14.9	194	2 Q8JFW5_BRARE	Q8jfw5 brachydanio
23	708	14.6	165	2 Q5GHL0_HUMAN	Q5ghl0 homo sapien
24	702	14.7	165	2 Q5XG96_HUMAN	Q5xg96 homo sapien
25	659.5	13.7	880	2 Q7Q1Y1_ANOGA	Q7q1y1 anopheles g
26	654	13.6	1022	2 Q54QA0_DICDI	Q54qa0 dictyosteli
27	639	13.3	143	2 Q8BL19_MOUSE	Q8bl19 mus musculu
28	616.5	12.8	895	2 Q8T9C0_DROME	Q8t9c0 drosophila
29	615.5	12.8	899	2 Q9V7B6_DROME	Q9v7b6 drosophila
30	607	12.6	923	2 Q941I6_ARATH	Q941i6 arabidopsis
31	595.5	12.4	893	2 Q76417_DROME	Q76417 drosophila

32	587.5	12.2	805	2	Q9TVL8_CABEL	Q9tv18 caenorhabdi
33	584	12.1	871	2	Q5ZJ94_CHICK	Q5zj94 gallus gall
34	581.5	12.1	903	2	Q755U7_ASHGO	Q755u7 ashbya goss
35	572.5	11.9	797	2	Q6OM36_CABER	Q6om36 caenorhabdi
36	569	11.8	862	1	PMS2_HUMAN	P54278 homo sapien
37	569	11.8	862	2	Q75MR2_HUMAN	Q75mr2 homo sapien
38	569	11.8	879	2	Q8N5Q6_HUMAN	Q8n5q6 homo sapien
39	568	11.8	862	2	Q52LH6_HUMAN	Q52lh6 homo sapien
40	564.5	11.7	907	2	Q6PPA0_CANGA	Q6ppa0 candida gla
41	563	11.7	923	2	Q69L72_ORISA	Q69l72 oryza sativ
42	561.5	11.7	893	2	Q6C6B8_YARLI	Q6c6b8 yarrowia li
43	561	11.7	908	2	Q8TGS0_YEAST	Q8tgs0 saccharomyc
44	559.5	11.6	1094	2	Q4XWC3_PLACH	Q4xwc3 plasmodium
45	558	11.6	908	2	Q8TG48_YEAST	Q8tg48 saccharomyc

ALIGNMENTS

RESULT 1

ID	PMS1_HUMAN	STANDARD;	PRT;	932 AA.
AC	P54277;			
DT	01-OCT-1996 (Rel. 34, Created)			
DT	01-OCT-1996 (Rel. 34, Last sequence update)			
DT	10-MAY-2005 (Rel. 47, Last annotation update)			
DE	PMS1 protein homolog 1 (DNA mismatch repair protein PMS1).			
GN	Name=PMS1; Synonym=SPMS1;			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;			
OC	Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	NUCLEOTIDE SEQUENCE.			
RC	TISSUE=Small intestine;			
RX	MEDLINE=94352394; PubMed=8072530; DOI=10.1038/371075a0;			
RA	Nicolaides N.C., Papadopoulos N., Liu B., Wei Y.-F., Carter K.C.,			
RA	Ruben S.M., Rosen C.A., Haseltine W.H., Fleischmann R.D., Fraser C.M.,			
RA	Adams M.D., Venter J.C., Dunlop M.G., Hamilton S.R., Petersen G.M.,			
RA	de la Chapelle A., Vogelstein B., Kinzler K.W.;			
RT	"Mutations of two PMS homologues in hereditary nonpolyposis colon cancer.";			
RL	Nature 371:75-80(1994).			
RN	[2]			
RP	NUCLEOTIDE SEQUENCE [GENOMIC DNA], AND VARIANTS GLN-27; LYS-202;			
RP	ARG-501; SER-632; ASP-720 AND HIS-793.			
RA	Rieder M.J., Livingston R.J., Daniels M.R., Chung M.-W.,			
RA	Miyamoto K.E., Nguyen C.P., Nguyen D.A., Poel C.L., Robertson P.D.,			
RA	Schackwitz W.S., Sherwood J.K., Witrak L.A., Nickerson D.A.;			
RT	"NIHES-SNP, environmental genome project, NIHES ES15478, Department of Genome Sciences, Seattle, WA (URL: http://egp.gs.washington.edu).";			
RT	Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.			
RL	[3]			
RP	VARIANTS HNPCC3 THR-394 AND ARG-501.			
RX	MEDLINE=9408236; PubMed=10480159; DOI=10.1007/s004390051067;			
RA	Wang Q., Lasset C., Desseigne F., Saurin J.-C., Maugard C., J.-F.,			
RA	Navarro C., Ruano E., Descos L., Trillet-Lenoir V., Bosset J.-F.,			
RT	Puisieux A.;			
RT	"Prevalence of germline mutations of hMLH1, hMSH2, hPMS1, and hMSH6 genes in 75 French kindreds with nonpolyposis colorectal cancer.";			
RL	Hum. Genet. 105:79-85(1999).			
CC	- - FUNCTION: Probably involved in the repair of mismatches in DNA.			
CC	- - SUBCELLULAR LOCATION: Nuclear (Potential).			
CC	- - DISEASE: Defects in PMS1 are the cause of hereditary non-polyposis colorectal cancer type 3 (HNPCC3) [MIM:600258]. Mutations in more than one gene locus can be involved alone or in combination in the production of the HNPCC phenotype (also called Lynch syndrome).			
CC	Most families with clinically recognized HNPCC have mutations in either MLH1 or MSH2 genes. HNPCC is an autosomal, dominantly inherited disease associated with marked increase in cancer susceptibility. It is characterized by a familial predisposition			

to early onset colorectal carcinoma (CRC) and extra-colonic cancers of the gastrointestinal, urological and female reproductive tracts. HNPCC is reported to be the most common form of inherited colorectal cancer in the Western world, and accounts for 15% of all colon cancers. Cancers in HNPCC originate within benign neoplastic polyps termed adenomas. Clinically, HNPCC is often divided into two subgroups. Type I: hereditary predisposition to colorectal cancer, a young age of onset, and carcinoma observed in the proximal colon. Type II: patients have an increased risk for cancers in certain tissues such as the uterus, ovary, breast, stomach, small intestine, skin, and larynx in addition to the colon. Diagnosis of classical HNPCC is based on the Amsterdam criteria: 3 or more relatives affected by colorectal cancer, one a first degree relative of the other two; 2 or more generation affected; 1 or more colorectal cancers presenting before 50 years of age; exclusion of hereditary polyposis syndromes. The term "suspected HNPCC" or "incomplete HNPCC" can be used to describe families who do not or only partially fulfill the Amsterdam criteria, but in whom a genetic basis for colon cancer is strongly suspected.

CC -!- SIMILARITY: Belongs to the DNA mismatch repair mutL/hexB family.
 CC -!- SIMILARITY: Contains 1 HMG box DNA-binding domain.
 CC -!- DATABASE: NAME=Hereditary non-polyposis colorectal cancer db;
 WWW="http://www.nfdht.nl/"

 *This Swiss-Prot entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use as long as its content is in no way modified and this statement is not removed.

 EMBL; U13695; AA63922.1; -; Genomic DNA.
 EMBL; AY267352; AA089079.1; -; Genomic DNA.
 PIR; S47597; S47597.
 HSSP; P54278; 1H7S.
 DR Ensembl; ENSG00000064933; Homo sapiens.
 DR HGNC; HGNC:9121; PMS1.
 DR MIM; 600258; -; C.nucleus; TAS.
 DR GO; GO:0005634; C.nucleus; TAS.
 DR GO; GO:0003677; F:DNA binding; TAS.
 DR GO; GO:0006298; P:mismatch repair; TAS.
 DR InterPro; IPR003594; ATP bd ATPase.
 DR InterPro; IPR002099; DNA_mis_repair.
 DR InterPro; IPR000910; HMG_12_Box.
 DR PANTHER; PTHR10073; DNA_mis_repair; 1.
 DR Pfam; PF01119; DNA_mis_repair; 1.
 DR Pfam; PF02518; HATPase_c; 1.
 DR Pfam; PF05055; HMG_box; 1.
 DR TIGRfam; TIGR00585; mutL; 1.
 DR PROSITE; PS00058; DNA_MISMATCH_REPAIR_1; 1.
 DR PROSITE; PS0118; HMG_BOX_2; 1.
 KW Anti-oncogene; Cell cycle; Disease mutation; DNA damage; DNA repair;
 KW Hereditary nonpolyposis colorectal cancer; Nuclear protein;
 KW Polymorphism.
 FT DNA_BIND 571 639 HMG box.
 FT VARIANT 27 27 E -> Q (in dbSNP:5742973).
 FT VARIANT 202 202 R -> K (in dbSNP:2066459).
 FT VARIANT 394 394 M -> T (in dbSNP:2066456).
 FT VARIANT 501 501 G -> R (in dbSNP:1145232).
 FT VARIANT 632 632 N -> S (in dbSNP:2066456).
 FT VARIANT 720 720 E -> D (in dbSNP:2066455).
 FT VARIANT 793 793 Y -> H (in dbSNP:1145234).
 FT SEQUENCE 932 AA; 105830 MW; EC4F402937B616DF CRC64;

Query Match	100.0%	Score 4812;	DB 1;	Length 932;
Best Local Similarity	100.0%	Pred. No. 1.2e-214;		
Matches 932;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	MKQLPAAVTRLLSSQIITSVVVVKELIENSLDAGATSDVKLENYGFDFKIEVDNGBG	60	
Db	1	MKQLPAAVTRLLSSQIITSVVVVKELIENSLDAGATSDVKLENYGFDFKIEVDNGBG	60	
QY	61	IKAVDAPVMAMKYTTSKINSHEDLENLTTCYGRGALGSIICIAEVLITTRTAADNFSTQ	120	
Db	61	IKAVDAPVMAMKYTTSKINSHEDLENLTTCYGRGALGSIICIAEVLITTRTAADNFSTQ	120	
QY	121	YVLDSGSHILSQKPSHLGGTTVTRALRFLKNLPVRKQFYSTAKKCKBIKTIQDMLMSFG	180	
Db	121	YVLDSGSHILSQKPSHLGGTTVTRALRFLKNLPVRKQFYSTAKKCKBIKTIQDMLMSFG	180	
QY	181	ILKPDRLRVFVHKAVIWKSRVSDHKALMSVLGTAVMNMNMFQYHSESOIYLSGFL	240	
Db	181	ILKPDRLRVFVHKAVIWKSRVSDHKALMSVLGTAVMNMNMFQYHSESOIYLSGFL	240	
QY	241	PKCADHSFTSLSTPERSFIFINRSPVHKDKILIRHHYNLCKLKESTRLYPVFLKID	300	
Db	241	PKCADHSFTSLSTPERSFIFINRSPVHKDKILIRHHYNLCKLKESTRLYPVFLKID	300	
QY	301	VPTADVVDNLTPDKSQVLLQNKESVLIENLMTTCYGLPSTNSYENNKTDVSAADIVL	360	
Db	301	VPTADVVDNLTPDKSQVLLQNKESVLIENLMTTCYGLPSTNSYENNKTDVSAADIVL	360	
QY	361	SKTATDVLFNKVRSSGKNYNDVTSVTPFQNDMHNDESGKNTDCLNHOISIGDFGYGH	420	
Db	361	SKTATDVLFNKVRSSGKNYNDVTSVTPFQNDMHNDESGKNTDCLNHOISIGDFGYGH	420	
QY	421	CSSEISNIDKNTKNAFQDISMSNSWENSQTEYSKTCFISSVKHTQSENGKNDHIDEGE	480	
Db	421	CSSEISNIDKNTKNAFQDISMSNSWENSQTEYSKTCFISSVKHTQSENGKNDHIDEGE	480	
QY	481	NEEEAGLENSSEISADEWSRGNILKNSVGENIEPVKILVPEKSLPCKVSNNNYPPEQWN	540	
Db	481	NEEEAGLENSSEISADEWSRGNILKNSVGENIEPVKILVPEKSLPCKVSNNNYPPEQWN	540	
QY	541	LNEDSCNKKSNVIDNKGKVTAYDILLNRRVKKPMSASALFVODHRPOFLIENPKTSLD	600	
Db	541	LNEDSCNKKSNVIDNKGKVTAYDILLNRRVKKPMSASALFVODHRPOFLIENPKTSLD	600	
QY	601	ATLQIEELWKLTSBEEKLYEKKATKOLERNYSOMKRAIEQESQMSLKDGRKKIKPTSAW	660	
Db	601	ATLQIEELWKLTSBEEKLYEKKATKOLERNYSOMKRAIEQESQMSLKDGRKKIKPTSAW	660	
QY	661	NLAQKHKLKTSLSNQPKLDELQSQIEKRRSQNIKWQIPFSMKNLKINFKQNKVDLEE	720	
Db	661	NLAQKHKLKTSLSNQPKLDELQSQIEKRRSQNIKWQIPFSMKNLKINFKQNKVDLEE	720	
QY	721	KDEPCLIHNLAPPDAMLMTKTEVMLNPPYRVEEALLFKLLENHKLPAEPLEKPTMLTE	780	
Db	721	KDEPCLIHNLAPPDAMLMTKTEVMLNPPYRVEEALLFKLLENHKLPAEPLEKPTMLTE	780	
QY	781	SLFNGSHYLDVLYKMTADDQRYSGSTYLSDBPLTANGFKIKLIPGVSIPTENYLEIEGMAN	840	
Db	781	SLFNGSHYLDVLYKMTADDQRYSGSTYLSDBPLTANGFKIKLIPGVSIPTENYLEIEGMAN	840	
QY	841	CLPFGYGVADLKEILNAILNRNAKEVEYECRPRKVISYLEGEAVRLSRQLPMYLSKEDIQDI	900	
Db	841	CLPFGYGVADLKEILNAILNRNAKEVEYECRPRKVISYLEGEAVRLSRQLPMYLSKEDIQDI	900	
QY	901	IYRMKHQFGNEIKECVHGRRPPFHHTYLPETT	932	
Db	901	IYRMKHQFGNEIKECVHGRRPPFHHTYLPETT	932	

RESULT 2

Q4VAL4 HUMAN

ID Q4VAL4 HUMAN PRELIMINARY; PRT; 920 AA.

AC Q4VAL4;

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model
Run on: December 21, 2005, 20:24:27 ; Search time 46 Seconds
(without alignments)
1949.435 Million cell updates/sec

Title: US-10-079-429A-4
Perfect score: 932
Sequence: 1 MKQLPAATVRLSSQIITS.....KECVHGRPFPHLTLVLETT 932

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 283416 seqs, 96216763 residues
Word size : 0

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : PIR 80:*
1: PIR1:*
2: PIR2:*
3: PIR3:*
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	932	100.0	932	2 S47597	mutL protein homolog
2	16	1.7	580	2 A72032	DNA mismatch repair
3	16	1.7	580	2 B86592	DNA mismatch repair
4	15	1.6	576	2 A71497	probable DNA mismatch
5	12	1.3	779	2 T01304	hypothetical prote
6	11	1.2	684	2 T50317	probable DNA mismatch
7	9	1.0	98	2 JC2403	PMS8 homolog
8	9	1.0	159	2 JC2401	PMS6 homolog
9	9	1.0	161	2 JC2402	PMS7 homolog
10	9	1.0	186	2 JC2400	PMS5 homolog
11	9	1.0	249	2 D96691	hypothetical prote
12	9	1.0	252	2 JC2399	PMS4 homolog
13	9	1.0	256	2 JC2398	PMS3 homolog
14	9	1.0	425	2 D70436	DNA mismatch repair
15	9	1.0	516	2 H72427	DNA mismatch repair
16	9	1.0	530	2 C90248	DNA topoisomerase
17	9	1.0	576	2 G81657	DNA mismatch repair
18	9	1.0	615	2 PH0853	methyl-directed mi
19	9	1.0	615	2 B91272	enzyme in methyl-d
20	9	1.0	615	2 B86113	DNA mismatch repair
21	9	1.0	618	2 AG1048	DNA mismatch repair
22	9	1.0	618	2 A33588	mismatch repair pr
23	9	1.0	635	2 AC0046	DNA mismatch repair
24	9	1.0	653	2 A82334	DNA mismatch repair
25	9	1.0	756	2 S43085	DNA mismatch repair
26	9	1.0	769	2 S54525	mismatch repair pr
27	9	1.0	779	2 T25389	hypothetical prote
28	9	1.0	862	2 S47598	mutL protein homolog
29	8	0.9	114	2 T16365	hypothetical prote

30 8 0.9 116 2 G84032 hypothetical prote
31 8 0.9 213 2 D69409 conserved hypothet
32 8 0.9 259 2 A86409 hypothetical prote
33 8 0.9 305 2 B84274 heme biosynthesis
34 8 0.9 336 2 A81409 probable periplasm
35 8 0.9 349 2 H64042 signal peptidase I
36 8 0.9 392 2 F96937 cell wall-associat
37 8 0.9 393 2 AG0091 probable flagellar
38 8 0.9 446 2 D71418 hypothetical prote
39 8 0.9 474 2 B86221 hypothetical prote
40 8 0.9 492 2 T00433 fumarate hydratase
41 8 0.9 532 2 H83993 two-component sens
42 8 0.9 563 2 AH2187 DNA mismatch repair
43 8 0.9 584 2 JQ1229 cellulase (EC 3.2.
44 8 0.9 629 2 E64046 mismatch repair pr
45 8 0.9 659 2 E84176 DNA mismatch repair

ALIGNMENTS

RESULT 1

S47597
mutL protein homolog - human
C:Species: Homo sapiens (man)
C:Date: 27-Jan-1995 #sequence_revision 27-Jan-1995 #text_change 09-Jul-2004
C:Accession: S47597
R:Nicolaides, N.C.; Papadopoulos, N.; Liu, B.; Wei, Y.F.; Carter, K.C.; Ruben, S.M.; R.
S.R.; Petersen, G.W.; de la Chapelle, A.; Vogelstein, B.; Kinzler, K.W.
Nature 371, 75-80, 1994
A:Title: Mutations of two PMS homologues in hereditary nonpolyposis colon cancer.
A:Reference number: S47597; MUID:94352394; PMID:8072530
A:Accession: S47597
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-932 <NIC>
A:Cross-references: UNIPROT:P54277; UNIPARC:UPI00000405F5; EMBL:UI3695; NID:g535512; P
C:Genetics:
A:Gene: GDB:PMS1; PMSL1
A:Cross-references: GDB:386403; OMIM:600258
A:Map position: 2q31-2q33
F:571-643/Domain: HMG box homology <HMG1>

Query Match	100.0%	Score 932;	DB 2;	Length 932;
Best Local Similarity	100.0%	Pred. No. 0;		
Matches 932;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	MKQLPAATVRLSSQIITSVVSVVKELIENSLDAGATSDVKLENYGDFDKIEVRNGEG	60	
DB	1	MKQLPAATVRLSSQIITSVVSVVKELIENSLDAGATSDVKLENYGDFDKIEVRNGEG	60	
QY	61	IKAVDAPVMAMKYTSKINSHEDLENLTYYGFRGALGSIICCAEVLITTRTADNFSTQ	120	
DB	61	IKAVDAPVMAMKYTSKINSHEDLENLTYYGFRGALGSIICCAEVLITTRTADNFSTQ	120	
QY	121	VYLDGSGHILSKPQSHLGQGTITVTAIRLFKNLPVRKQFYSTAKCKDEIKKIQDLMSFG	180	
DB	121	VYLDGSGHILSKPQSHLGQGTITVTAIRLFKNLPVRKQFYSTAKCKDEIKKIQDLMSFG	180	
QY	181	ILKPLDIRIVFVNKAVIQKSRVSDHKMALMSVLGTAVVNNMESFOYHSEESQIYLSGFL	240	
DB	181	ILKPLDIRIVFVNKAVIQKSRVSDHKMALMSVLGTAVVNNMESFOYHSEESQIYLSGFL	240	
QY	241	PKCDADHSFTSLSTPERSIFINSPVHOKDILKIRHHYNLCKLESTRLYPVFFLKID	300	
DB	241	PKCDADHSFTSLSTPERSIFINSPVHOKDILKIRHHYNLCKLESTRLYPVFFLKID	300	
QY	301	VPTADVNNLTDPKQVLLQNKESVLIALENLMTTCYGPSTNSYENNKTDVSAADIVL	360	
DB	301	VPTADVNNLTDPKQVLLQNKESVLIALENLMTTCYGPSTNSYENNKTDVSAADIVL	360	
QY	361	SKTAEITDVLFNKVESGKNYSNVDTSVIPFQNDMNDGSKQNTDCLNHQISIGDGYGH	420	

Db 361 SKTATDVLVFNKVESGKNYSNVDTSVIPFQDMHNDBSGKVTDDCLNHQISIGDFGYCH 420

Qy 421 CSSEISNIDKNTKNAFQDIMSNSWNSQTSYKTCFTISSVKHTQSENGNKHIDESGE 480

Db 421 CSSEISNIDKNTKNAFQDIMSNSWNSQTSYKTCFTISSVKHTQSENGNKHIDESGE 480

Qy 481 NEEERAGLENSSEISADEWSRGNLKNSVGENIEPVKILVPEKSLPCKVSNNNYPPEOWN 540

Db 481 NEEERAGLENSSEISADEWSRGNLKNSVGENIEPVKILVPEKSLPCKVSNNNYPPEOWN 540

Qy 541 LNEDESCNKSNSVINDKSGKVTAYDILLSNRVVIKPPMSASALFVQDHRPQFLIENPKTSLED 600

Db 541 LNEDESCNKSNSVINDKSGKVTAYDILLSNRVVIKPPMSASALFVQDHRPQFLIENPKTSLED 600

Qy 601 ATLQIEELWKTLSSEKLYEKATKDLRYNSOMKRAIEQESOMSLKDGKRKIKPTSAW 660

Db 601 ATLQIEELWKTLSSEKLYEKATKDLRYNSOMKRAIEQESOMSLKDGKRKIKPTSAW 660

Qy 661 NLAQKHKLTSLSNPKDELLOSOIEKRRSONIKWQIIPFSMKNLKINFKNKONKVDLEE 720

Db 661 NLAQKHKLTSLSNPKDELLOSOIEKRRSONIKWQIIPFSMKNLKINFKNKONKVDLEE 720

Qy 721 KDBPCLHNLRPDPAWMTSKTEVMLNPNRYVEEALLFKRLLENHKLPAEPLKPTMLTE 780

Db 721 KDBPCLHNLRPDPAWMTSKTEVMLNPNRYVEEALLFKRLLENHKLPAEPLKPTMLTE 780

Qy 781 SLFNGSHYLDVLYKMTADQRYSGSTYLSDPRLTANGFKILIPGVSIITENYLETEGMAN 840

Db 781 SLFNGSHYLDVLYKMTADQRYSGSTYLSDPRLTANGFKILIPGVSIITENYLETEGMAN 840

Qy 841 CLPFFGVADLKEITANILNRNAKEVCEPRKVIISYLEGEAVRLSRQLPMYLSKEDIQDI 900

Db 841 CLPFFGVADLKEITANILNRNAKEVCEPRKVIISYLEGEAVRLSRQLPMYLSKEDIQDI 900

Qy 901 IYRMKHQFNGEIKECVHGRRPFFHHLTYLPETT 932

Db 901 IYRMKHQFNGEIKECVHGRRPFFHHLTYLPETT 932

RESULT 2

A72032

DNA mismatch repair protein mutL CP1059 [imported] - Chlamydomophila pneumoniae (strains C

C:Species: Chlamydomophila pneumoniae, Chlamydia pneumoniae

C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 09-Jul-2004

C:Accession: A72032; F81506

R:Kallman, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood, J.; Nature Genet. 21, 385-389, 1999

A:Title: Comparative Genomes of Chlamydia pneumoniae and C. trachomatis.

A:Reference number: A72000; MUID:99206606; PMID:10192388

A:Accession: A72032

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-580 <ARN>

A:Cross-references: UNIPROT:Q92794; UNIPARC:UPI000012FA3C; GB:AE001662; GB:AE001363; NID

A:Experimental source: strain CWL029

R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey, C.; Dodson, R.; Gwinn, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzberg, Nucleic Acids Res. 28, 1397-1406, 2000

A:Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39.

A:Reference number: A81500; MUID:20150255; PMID:10684935

A:Accession: F81506

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-580 <REA>

A:Cross-references: UNIPARC:UPI000012FA3C; GB:AE002263; GB:AE002161; NID:g7189971; PIDN:

A:Experimental source: strain AR39, HL cells

C:Genetics:

A:Gene: mutL; CP1059

C:Superfamily: mismatch repair protein hexB

Query Match 1.7%; Score 16; DB 2; Length 580;

Best Local Similarity 100.0%; Pred. No. 1.1e-07;

Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 22 VSVVKELIENSLDAGA 37

Db 28 VSVVKELIENSLDAGA 43

RESULT 3

B86592

DNA mismatch repair [imported] - Chlamydomophila pneumoniae (strain J138)

C:Species: Chlamydomophila pneumoniae, Chlamydia pneumoniae

C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 09-Jul-2004

C:Accession: B86592

R:Shirai, M.; Hirakawa, H.; Kimoto, M.; Tabuchi, M.; Kishi, F.; Ouchi, K.; Shiba, T.; Nucleic Acids Res. 28, 2311-2314, 2000

A:Title: Comparison of whole genome sequences of chlamydia pneumoniae J138.

A:Reference number: A86491; MUID:20330349; PMID:10871362

A:Accession: B86592

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-580 <STO>

A:Cross-references: UNIPROT:Q92794; UNIPARC:UPI000012FA3C; GB:BA000008; NID:g8979186;

A:Experimental source: strain J138

C:Genetics:

A:Gene: mutL

C:Superfamily: mismatch repair protein hexB

Query Match 1.7%; Score 16; DB 2; Length 580;

Best Local Similarity 100.0%; Pred. No. 1.1e-07;

Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 22 VSVVKELIENSLDAGA 37

Db 28 VSVVKELIENSLDAGA 43

RESULT 4

A71497

probable DNA mismatch repair - Chlamydia trachomatis (serotype D, strain UW3/Cx)

C:Species: Chlamydia trachomatis

C:Date: 13-Sep-1998 #sequence_revision 13-Sep-1998 #text_change 09-Jul-2004

C:Accession: A71497

R:Stephens, R.S.; Kallman, S.; Lammel, C.J.; Fan, J.; Marathe, R.; Aravind, L.; Mitchell Science 282, 754-759, 1998

A:Title: Genome sequence of an obligate intracellular pathogen of humans: Chlamydia tr

A:Reference number: A71570; MUID:99000809; PMID:9784136

A:Accession: A71497

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-576 <ARN>

A:Cross-references: UNIPROT:O84579; UNIPARC:UPI000012FA3E; GB:AE001328; GB:AE001273; N

A:Experimental source: serotype D, strain UW-3/Cx

C:Genetics:

A:Gene: mutL

C:Superfamily: mismatch repair protein hexB

Query Match 1.6%; Score 15; DB 2; Length 576;

Best Local Similarity 100.0%; Pred. No. 1.2e-06;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 23 SVVKELIENSLDAGA 37

Db 30 SVVKELIENSLDAGA 44

RESULT 5

T01304

hypothetical protein T14P8.6 - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cross)

C:Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 09-Jul-2004

C:Accession: T01304

R:Kalicki, J.; Elliott, G.; Cloud, J. submitted to the EMBL Data Library, May 1998

A:Description: The sequence of A. thaliana T14P8.

A:Reference number: Z14290
A:Accession: T01304
A>Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-779 <KAL>
A:Cross-references: UNIPROT:O81287; UNIPARC:UPI0000040685; EMBL:AF069298; NID:g3193282;
A:Experimental source: cultivar Columbia
C:Genetics:
A:Map position: 4
A:Introns: 94/3; 202/3; 254/3; 562/3; 585/3; 610/2; 632/2; 667/1; 727/3
A>Note: T14P8.6
C:Superfamily: DNA mismatch repair protein

Query Match 1.3%; Score 12; DB 2; Length 779;
Best Local Similarity 100.0%; Pred. No. 0.0021;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 86 NLTYYGFRGEAL 97
|||||
Db 113 NLTYYGFRGEAL 124
|||||

RESULT 6
T50317
probable DNA mismatch repair protein, MLH1 homolog [imported] - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C:Date: 09-Jun-2000 #sequence_revision 09-Jun-2000 #text_change 09-Jul-2004
C:Accession: T50317
R:McDougall, R.C.; Rajandream, M.A.; Barrell, B.G.; Cadieu, V.; Lelaure, V.; Galibert, F.
submitted to the EMBL Data Library, January 2000
A:Reference number: Z25061
A:Accession: T50317
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-684 <MCD>
A:Cross-references: UNIPROT:Q9P7W6; UNIPARC:UPI000006AC8D; EMBL:AL136536; PIDN:CAB66448.
A:Experimental source: strain 972h(-); cosmid c1703
C:Genetics:
A:Gene: SPDB:SPBC1703.04
A:Map position: 2
A:Introns: 24/3; 70/3; 128/2
C:Superfamily: DNA mismatch repair protein, Mlh1 type

Query Match 1.2%; Score 11; DB 2; Length 684;
Best Local Similarity 100.0%; Pred. No. 0.02;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 26 KELIENSLDAG 36
|||||
Db 34 KELIENSLDAG 44
|||||

RESULT 7
JC2403
PMS8 homolog mismatch repair protein - human
C:Species: Homo sapiens (man)
C:Date: 24-Feb-1995 #sequence_revision 24-Feb-1995 #text_change 09-Jul-2004
C:Accession: JC2403
R:Horii, A.; Han, H.J.; Sasaki, S.; Shimada, M.; Nakamura, Y.
Biochem. Biophys. Res. Commun. 204, 1257-1264, 1994
A:Title: Cloning, characterization and chromosomal assignment of the human genes homolog
A:Reference number: JC2398; MUID:95071462; PMID:7980603
A:Accession: JC2403
A:Molecule type: DNA
A:Residues: 1-98 <HOR>
A:Cross-references: UNIPROT:Q16590; UNIPARC:UPI0000073FED; DBJ:D38440; NID:g600595; PIDN:CAB66448.
C:Genetics:
A:Gene: GDB:PMS2L6; PMS8
A:Cross-references: GDB:437147
A:Map position: 7q11.23-7q22
C:Keywords: DNA repair

Query Match 1.0%; Score 9; DB 2; Length 98;

Best Local Similarity 100.0%; Pred. No. 0.44;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 30 ENSLDAGAT 38
|||||
Db 48 ENSLDAGAT 56
|||||

RESULT 8
JC2401
PMS6 homolog mismatch repair protein - human
C:Species: Homo sapiens (man)
C:Date: 24-Feb-1995 #sequence_revision 24-Feb-1995 #text_change 09-Jul-2004
C:Accession: JC2401
R:Horii, A.; Han, H.J.; Sasaki, S.; Shimada, M.; Nakamura, Y.
Biochem. Biophys. Res. Commun. 204, 1257-1264, 1994
A:Title: Cloning, characterization and chromosomal assignment of the human genes homolog
A:Reference number: JC2398; MUID:95071462; PMID:7980603
A:Accession: JC2401
A:Molecule type: DNA
A:Residues: 1-159 <HOR>
A:Cross-references: UNIPROT:Q16544; UNIPARC:UPI0000073EE1; DBJ:D38438; NID:g600593; P
C:Genetics:
A:Gene: GDB:PMS2L4; PMS6
A:Cross-references: GDB:437145
A:Map position: 7q11.23-7q22
C:Keywords: DNA repair

Query Match 1.0%; Score 9; DB 2; Length 159;
Best Local Similarity 100.0%; Pred. No. 0.68;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 30 ENSLDAGAT 38
|||||
Db 49 ENSLDAGAT 57
|||||

RESULT 9
JC2402
PMS7 homolog mismatch repair protein - human
C:Species: Homo sapiens (man)
C:Date: 24-Feb-1995 #sequence_revision 24-Feb-1995 #text_change 09-Jul-2004
C:Accession: JC2402
R:Horii, A.; Han, H.J.; Sasaki, S.; Shimada, M.; Nakamura, Y.
Biochem. Biophys. Res. Commun. 204, 1257-1264, 1994
A:Title: Cloning, characterization and chromosomal assignment of the human genes homolog
A:Reference number: JC2398; MUID:95071462; PMID:7980603
A:Accession: JC2402
A:Molecule type: DNA
A:Residues: 1-161 <HOR>
A:Cross-references: UNIPROT:Q16673; UNIPARC:UPI000017C30B; DBJ:D38439
C:Genetics:
A:Gene: GDB:PMS2L5; PMS7
A:Cross-references: GDB:437146
A:Map position: 7q11.23-7q22
C:Keywords: DNA repair

Query Match 1.0%; Score 9; DB 2; Length 161;
Best Local Similarity 100.0%; Pred. No. 0.69;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 30 ENSLDAGAT 38
|||||
Db 33 ENSLDAGAT 41
|||||

RESULT 10
JC2400
PMS5 homolog mismatch repair protein - human
C:Species: Homo sapiens (man)
C:Date: 24-Feb-1995 #sequence_revision 24-Feb-1995 #text_change 09-Jul-2004
C:Accession: JC2400
R:Horii, A.; Han, H.J.; Sasaki, S.; Shimada, M.; Nakamura, Y.

Biochem. Biophys. Res. Commun. 204, 1257-1264, 1994
A:Title: Cloning, characterization and chromosomal assignment of the human genes homolog
A:Reference number: JC2398; MUID:95071462; PMID:7980603
A:Accession: JC2400
A:Molecule type: DNA
A:Residues: 1-186 <HOR>
A:Cross-references: UNIPROT:Q16603; UNIPARC:UPI0000073EE3; DDBJ:D38437; NID:g600592; PID
C:Genetics:
A:Gene: GDB:PMS2L3; PMS5
A:Cross-references: GDB:437144
A:Map position: 7q11.23-7q22
C:Keywords: DNA repair

Query Match 1.0%; Score 9; DB 2; Length 186;
Best Local Similarity 100.0%; Pred. No. 0.78;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 30 ENSLDAGAT 38
|||||
Db 38 ENSLDAGAT 46
|||||

RESULT 11
D96691
hypothetical protein T1217.1 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana [mouse-ear cress]
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 09-Jul-2004
C:Accession: D96691
R:Theologos, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Mafti, R.; Marziani,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719; PMID:11130712
A:Accession: D96691
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-249 <STO>
A:Cross-references: UNIPROT:Q9C557; UNIPARC:UPI000009D93C; GB:AE005173; NID:gl1054575; F
C:Genetics:
A:Gene: T1217.1
A:Map position: 1

Query Match 1.0%; Score 9; DB 2; Length 249;
Best Local Similarity 100.0%; Pred. No. 1;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 759 KRLENNHKL 767
|||||
Db 24 KRLENNHKL 32
|||||

RESULT 12
JC2399
PMS4 homolog mismatch repair protein - human
C:Species: Homo sapiens (man)
C:Date: 24-Feb-1995 #sequence_revision 24-Feb-1995 #text_change 09-Jul-2004
C:Accession: JC2399
R:Horii, A.; Han, H.J.; Sasaki, S.; Shimada, M.; Nakamura, Y.
Biochem. Biophys. Res. Commun. 204, 1257-1264, 1994
A:Title: Cloning, characterization and chromosomal assignment of the human genes homolog
A:Reference number: JC2398; MUID:95071462; PMID:7980603
A:Accession: JC2399
A:Molecule type: DNA
A:Residues: 1-252 <HOR>
A:Cross-references: UNIPROT:Q16530; UNIPARC:UPI000017C30A; DDBJ:D38436
C:Genetics:
A:Gene: GDB:PMS2L2; PMS4

A:Cross-references: GDB:437143
A:Map position: 7q11.23-7q22

Query Match 1.0%; Score 9; DB 2; Length 252;
Best Local Similarity 100.0%; Pred. No. 1;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 30 ENSLDAGAT 38
|||||
Db 56 ENSLDAGAT 64
|||||

RESULT 13
JC2398
PMS3 homolog mismatch repair protein - human
C:Species: Homo sapiens (man)
C:Date: 24-Feb-1995 #sequence_revision 24-Feb-1995 #text_change 09-Jul-2004
C:Accession: JC2398
R:Horii, A.; Han, H.J.; Sasaki, S.; Shimada, M.; Nakamura, Y.
Biochem. Biophys. Res. Commun. 204, 1257-1264, 1994
A:Title: Cloning, characterization and chromosomal assignment of the human genes homolog
A:Reference number: JC2398; MUID:95071462; PMID:7980603
A:Accession: JC2398
A:Molecule type: DNA
A:Residues: 1-256 <HOR>
A:Cross-references: UNIPROT:Q16530; UNIPARC:UPI000017C309; DDBJ:D38435
C:Genetics:
A:Gene: GDB:PMS2L1; PMS3
A:Cross-references: GDB:437142
A:Map position: 7q11.23-7q22
C:Keywords: DNA repair

Query Match 1.0%; Score 9; DB 2; Length 256;
Best Local Similarity 100.0%; Pred. No. 1;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 30 ENSLDAGAT 38
|||||
Db 60 ENSLDAGAT 68
|||||

RESULT 14
D70436
DNA mismatch repair protein MutL - Aquifex aeolicus
C:Species: Aquifex aeolicus
C:Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 09-Jul-2004
C:Accession: D70436
R:Decker, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.;
V.
Nature 392, 353-358, 1998
A:Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.
A:Reference number: A70300; MUID:98196666; PMID:9537320
A:Accession: D70436
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-425 <AQF>
A:Cross-references: UNIPROT:O67518; UNIPARC:UPI0000056665; GB:AE000746; NID:g2983925;
A:Experimental source: strain VFS
C:Genetics:
A:Gene: mutL

Query Match 1.0%; Score 9; DB 2; Length 425;
Best Local Similarity 100.0%; Pred. No. 1.6;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 89 TYGFRGEAL 97
|||||
Db 91 TYGFRGEAL 99
|||||

RESULT 15
H72427
DNA mismatch repair protein - Thermotoga maritima (strain MSB8)

C:Species: Thermotoga maritima
C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jul-2000
C:Accession: H72427
R:Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hickey
Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.;
C.M.
Nature 399, 323-329, 1999
A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome seq
A:Reference number: A72200; MUID:99287316; PMID:10360571
A:Accession: H72427
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-516 <ARN>
A:Cross-references: UNIPARC:UPI0000166105; GB:AE001690; GB:AE000512; NID:g4980496; PIDN:
A:Experimental source: strain MSB8
C:Genetics:
A:Gene: TM0022
C:Superfamily: mismatch repair protein hexB
Query Match 1.0%; Score 9; DB 2; Length 516;
Best Local Similarity 100.0%; Pred. No. 1.9;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 89 TYGFRGEAL 97
Db 97 TYGFRGEAL 105
Search completed: December 21, 2005, 20:33:11
Job time : 48 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 21, 2005, 20:21:32 ; Search time 240 Seconds
(without alignments)
2739.802 Million cell updates/sec

Title: US-10-079-429A-4
Perfect score: 932
Sequence: 1 MKQLPAATVRLSSQIITS.....KECVHGRPFPHLTYLPETT 932

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 2166443 seqs, 705528306 residues

Word size: 0

Total number of hits satisfying chosen parameters: 2166443

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database: Uniprot 05.80.*
1: uniprot_sprot.*
2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	932	100.0	932	1 PMS1_HUMAN	P54277 homo sapien
2	920	98.7	920	2 Q4VAL4_HUMAN	Q4val4 homo sapien
3	899	75.0	893	2 Q5FBZ3_HUMAN	Q5fbz3 homo sapien
4	618	66.3	669	2 Q4VAL5_HUMAN	Q4val5 homo sapien
5	618	66.3	770	2 Q5FBZ8_HUMAN	Q5fbz8 homo sapien
6	463	49.7	667	2 Q5FBZ9_HUMAN	Q5fbz9 homo sapien
7	385	41.3	555	2 Q5FBZ6_HUMAN	Q5fbz6 homo sapien
8	234	25.1	234	2 Q5FBZ2_HUMAN	Q5fbz2 homo sapien
9	194	20.8	195	2 Q5FBZ1_HUMAN	Q5fbz1 homo sapien
10	194	20.8	196	2 Q5FBZ5_HUMAN	Q5fbz5 homo sapien
11	194	20.8	248	2 Q5FBZ4_HUMAN	Q5fbz4 homo sapien
12	161	17.3	163	2 Q68DF0_HUMAN	Q68df0 homo sapien
13	140	15.0	165	2 Q5XG96_HUMAN	Q5xg96 homo sapien
14	140	15.0	166	2 Q96HLO_HUMAN	Q96hlo homo sapien
15	137	14.7	931	2 Q5R904_PONPY	Q5r904 pongo pygma
16	44	4.7	47	2 Q5FBZ7_HUMAN	Q5fbz7 homo sapien
17	30	3.2	917	2 Q8K119_MOUSE	Q8k119 mus musculus
18	29	3.1	919	2 Q6P7D0_RAT	Q6p7d0 rattus norv
19	23	2.5	143	2 Q8BLI9_MOUSE	Q8bli9 mus musculus
20	18	1.9	916	2 Q5ZKT5_CHICK	Q5zkt5 gallus gall
21	16	1.7	580	1 MUTL_CHLPN	Q92794 chlamydia p
22	15	1.6	342	2 Q6ALT0_DESPS	Q6alt0 desulfotale
23	15	1.6	576	1 MUTL_CHLTR	Q84579 chlamydia t
24	15	1.6	1474	2 Q7QNV7_GIALA	Q7qnv7 giardia lam
25	15	1.6	1474	2 Q6WDAL_GIALA	Q6wdal giardia lam
26	14	1.5	854	2 Q4RTJ3_TETNG	Q4rtj3 tetraodon n
27	14	1.5	925	2 Q7ZXV9_XENLA	Q7zxv9 xenopus lae
28	14	1.5	928	2 Q5FVX9_XENTR	Q5fvx9 xenopus tro
29	13	1.4	194	2 Q8JFW5_BRARE	Q8jfw5 brachydanio
30	13	1.4	372	2 Q7SKD5_BRARE	Q7skd5 brachydanio
31	13	1.4	724	2 Q5JN46_ORYSA	Q5jn46 oryza sativ

32 13 1.4 831 2 Q4P3V5_USTMA Q4p3v5 ustilago ma
33 13 1.4 896 2 Q5JFR9_BRARE Q5jfr9 brachydanio
34 13 1.4 1111 2 Q52FC0_MAGGR Q52fc0 magnaporthe
35 12 1.3 426 1 MUTL_AQUPY P70754 aquifex pyr
36 12 1.3 733 2 Q8GY98_ARATH Q8gy98 arabidopsis
37 12 1.3 779 2 Q81287_ARATH Q81287 arabidopsis
38 12 1.3 923 2 Q941I6_ARATH Q941i6 arabidopsis
39 11 1.2 643 2 Q8RG56_FUSNN Q8rg56 fusobacteri
40 11 1.2 675 2 Q7P5M3_FUSNV Q7p5m3 fusobacteri
41 11 1.2 684 1 MLH1_SCHPO Q9p7w6 schizosacch
42 10 1.1 118 2 Q834Z3_ENTFA Q834z3 enterococcu
43 10 1.1 574 2 Q8JCM9_COXBU Q8jcm9 coxiella bu
44 10 1.1 624 1 MUTL_CHLTE Q8kax3 chloretellu
45 10 1.1 893 2 Q76417_DROME Q76417 drosophila

ALIGNMENTS

RESULT 1
PMS1_HUMAN STANDARD; PRT; 932 AA.
AC P54277;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 10-MAY-2005 (Rel. 47, Last annotation update)
DE PMS1 protein homolog 1 (DNA mismatch repair protein PMS1).
GN Name=PMS1; Synonyms=PMS1L;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC Homo.
OX NCBI_TaxID=9606;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=call bladder;
RX MEDLINE=94352394; PubMed=8072530; DOI=10.1038/371075a0;
RA Nicolaides N.C., Papadopoulos N., Liu B., Wei Y.-F., Carter K.C.,
RA Ruben S.M., Rosen C.A., Haseltine W.H., Fleischmann R.D., Fraser C.M.,
RA Adams M.D., Venter J.C., Dunlop M.G., Hamilton S.R., Petersen G.M.,
RA de la Chapelle A., Vogelstein B., Kinzler K.W.;
RT "Mutations of two PMS homologues in hereditary nonpolyposis colon
cancer.";
RL Nature 371:75-80 (1994).
RN [2]
RP NUCLEOTIDE SEQUENCE [GENOMIC DNA] AND VARIANTS GLN-27; LYS-202;
RP ARG-501; SER-532; ASP-720 AND HIS-793.
RA Rieder M.J., Livingston R.J., Daniels M.R., Chung M.-W.,
RA Miyamoto K.S., Nguyen C.P., Nguyen D.A., Poel C.L., Robertson P.D.,
RA Schackwitz W.S., Sherwood J.K., Witrak L.A., Nickerson D.A.;
RT "NIHES-SNPs, environmental genome project, NIHES ES15478, Department
of Genome Sciences, Seattle, WA (URL: http://egp.gs.washington.edu).";
RL Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.
RN [3]
RP VARIANTS HNPCC3 THR-394 AND ARG-501.
RX MEDLINE=99408236; PubMed=10480359; DOI=10.1007/s004390051067;
RA Wang Q., Lassot C., Desseigne F., Saurin J.-C., Maugard C.,
RA Navarro C., Ruano E., Descos L., Trillet-Lenoir V., Bosset J.-F.,
RA Puisieux A.;
RT "Prevalence of germline mutations of hMLH1, hMSH2, hPMS1, hPMS2, and
hMSH6 genes in 75 French kindreds with nonpolyposis colorectal
cancer.";
RL Hum. Genet. 105:79-85 (1999).
RN [4]
CC -!- FUNCTION: Probably involved in the repair of mismatches in DNA.
CC -!- SUBCELLULAR LOCATION: Nuclear (Potential).
CC -!- DISEASE: Defects in PMS1 are the cause of hereditary non-polyposis
CC colorectal cancer type 3 (HNPCC3) [MIM:600258]. Mutations in more
CC than one gene locus can be involved alone or in combination in the
CC production of the HNPCC phenotype (also called Lynch syndrome).
CC Most families with clinically recognized HNPCC have mutations in
CC either MLH1 or MSH2 genes. HNPCC is an autosomal, dominantly
CC inherited disease associated with marked increase in cancer
CC susceptibility. It is characterized by a familial predisposition

to early onset colorectal carcinoma (CRC) and extra-colonic cancers of the gastrointestinal, urological and female reproductive tracts. HNPCC is reported to be the most common form of inherited colorectal cancer in the western world, and accounts for 15% of all colon cancers. Cancers in HNPCC originate within benign neoplastic polyps termed adenomas. Clinically, HNPCC is often divided into two subgroups. Type I: hereditary predisposition to colorectal cancer, a young age of onset, and carcinoma observed in the proximal colon. Type II: patients have an increased risk for cancers in certain tissues such as the uterus, ovary, breast, stomach, small intestine, skin, and larynx in addition to the colon. Diagnosis of classical HNPCC is based on the Amsterdam criteria: 3 or more relatives affected by colorectal cancer, one a first degree relative of the other two; 2 or more generation affected; 1 or more colorectal cancers presenting before 50 years of age; exclusion of hereditary polyposis syndromes. The term "suspected HNPCC" or "incomplete HNPCC" can be used to describe families who do not or only partially fulfill the Amsterdam criteria, but in whom a genetic basis for colon cancer is strongly suspected.

--- SIMILARITY: Belongs to the DNA mismatch repair mutl/hexB family. ---
 -!- SIMILARITY: Contains 1 HMG box DNA-binding domain.
 -!- DATABASE: NAME=Hereditary non-polyposis colorectal cancer db; WWW="http://www.nfdht.nl/".

 This Swiss-Prot entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use as long as its content is in no way modified and this statement is not removed.

 EMBL; U13695; AAG31922.1; -; Genomic DNA.
 EMBL; AX267352; AAO89079.1; -; Genomic DNA.
 PIR; S47597; S47597.
 HSP; P54278; 1H7S.
 Ensembl; ENSG0000064933; Homo sapiens.
 HGNC; HGNC:9121; PMS1.
 MIM; 600258; -; Cnucleus; TAS.
 GO; GO:0005634; C:nucleus; TAS.
 GO; GO:0003677; F:DNA binding; TAS.
 GO; GO:006298; P:mismatch repair; TAS.
 InterPro; IPR003594; ATP bd ATPase.
 InterPro; IPR002099; DNA_mis repair.
 InterPro; IPR000910; HMG_12_Box.
 PANTHER; PTHR10073; DNA_mis repair; 1.
 Pfam; PF02518; HATPase_c; 1.
 Pfam; PF02518; HATPase_c; 1.
 TIGRFAMs; TIGR00585; mutl; 1.
 PROSITE; PS00058; DNA_MISMATCH_REPAIR_1; 1.
 PROSITE; PS00118; HMG_Box_2; 1.
 Anti-oncogene; Cell cycle; Disease mutation; DNA damage; DNA repair; Hereditary nonpolyposis colorectal cancer; Nuclear protein; Polymorphism.

FT DNA BIND 571 639 HMG box.
 FT VARIANT 27 27 E -> Q (in dbSNP:5742973).
 FT VARIANT 202 202 R -> K (in dbSNP:2066459).
 FT VARIANT 394 394 M -> T (in dbSNP:2066455).
 FT VARIANT 501 501 G -> R (in dbSNP:1145231).
 FT VARIANT 501 501 G -> R (in dbSNP:1145231).
 FT VARIANT 632 632 N -> S (in dbSNP:2066456).
 FT VARIANT 720 720 E -> D (in dbSNP:2066455).
 FT VARIANT 793 793 Y -> H (in dbSNP:1145234).
 FT SEQUENCE 932 AA; 105830 MW; EC4F402937B616DF CRC64;

Query Match	100.0%	Score 932;	DB 1;	Length 932;
Best Local Similarity	100.0%	Pred. No. 0;		
Matches 932;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY 1	MQLPAAVRLSSQITTSVVVKELIENSLDAGATSVDKLENYGDKIEVRDNGS	60		
DB 1	MQLPAAVRLSSQITTSVVVKELIENSLDAGATSVDKLENYGDKIEVRDNGS	60		
QY 61	IKAVDAPVAMKYITSKINSHEDLENLTYYGRGALGSCICIAEVLITRTAADNFSTQ	120		
DB 61	IKAVDAPVAMKYITSKINSHEDLENLTYYGRGALGSCICIAEVLITRTAADNFSTQ	120		
QY 121	VYLDGSHILSKPQSHLGGTTVTAIRLFPKNI PVKQFYSTAKCKDEIKKIQDLLMSFG	180		
DB 121	VYLDGSHILSKPQSHLGGTTVTAIRLFPKNI PVKQFYSTAKCKDEIKKIQDLLMSFG	180		
QY 181	ILKPLRLVFNHKAIVQKSRVSDHKMALMSVLGTAVNNMESFYHSEESQIYLSGFL	240		
DB 181	ILKPLRLVFNHKAIVQKSRVSDHKMALMSVLGTAVNNMESFYHSEESQIYLSGFL	240		
QY 241	PKCDADHFTSLSTPERSGFIFINSRPVHQKILKIRHHYNLCKLKESTRLYPVFLKID	300		
DB 241	PKCDADHFTSLSTPERSGFIFINSRPVHQKILKIRHHYNLCKLKESTRLYPVFLKID	300		
QY 301	VPTADVNLTPDKSQVLLQNKESVLI AL ENLMTTCYGLPSTNSYENNKTDVSAADIVL	360		
DB 301	VPTADVNLTPDKSQVLLQNKESVLI AL ENLMTTCYGLPSTNSYENNKTDVSAADIVL	360		
QY 361	SKTATDVLFNKVESGKNYSNVDTSVIPFQNDMNDSESKNTDCLNHOISIGDFGYGH	420		
DB 361	SKTATDVLFNKVESGKNYSNVDTSVIPFQNDMNDSESKNTDCLNHOISIGDFGYGH	420		
QY 421	CSSEISNIDKNTKNAFQDISMSNVSWENSQTESYKTCFISSVKHTQSENGKNDHIDSEGE	480		
DB 421	CSSEISNIDKNTKNAFQDISMSNVSWENSQTESYKTCFISSVKHTQSENGKNDHIDSEGE	480		
QY 481	NEEEAGLENSSEISADEWSRGNILKNSVGENTEPVKILVPEKSLPCKVNNNYPPEQWN	540		
DB 481	NEEEAGLENSSEISADEWSRGNILKNSVGENTEPVKILVPEKSLPCKVNNNYPPEQWN	540		
QY 541	LNEDSCNKSNI DNKSGKVATAYDLLSNRVIKKPMSASALFVQDHRPQFLIENPKTSLED	600		
DB 541	LNEDSCNKSNI DNKSGKVATAYDLLSNRVIKKPMSASALFVQDHRPQFLIENPKTSLED	600		
QY 601	ATLQTEELWKLTSBEEKLYEEKATKDLERYNSQMKRAIEQESQMSLKDGRKKIKPTSAW	660		
DB 601	ATLQTEELWKLTSBEEKLYEEKATKDLERYNSQMKRAIEQESQMSLKDGRKKIKPTSAW	660		
QY 661	NLAQKHKLKTSLSNQPKDELLOSOQIEKRRSONIKMVQIPFPMKNLKNFKKQNKVDLEE	720		
DB 661	NLAQKHKLKTSLSNQPKDELLOSOQIEKRRSONIKMVQIPFPMKNLKNFKKQNKVDLEE	720		
QY 721	KDEPCLIHNLRPDAMLTKTEVMLNPNRYVEALLPKRLLENHKLPAEPLKPIMLTE	780		
DB 721	KDEPCLIHNLRPDAMLTKTEVMLNPNRYVEALLPKRLLENHKLPAEPLKPIMLTE	780		
QY 781	SLFNGSHYLDVLYKMTADDQRYSGSTYLSDPRLTANGFKILIPGVSI TENYLEIEGMAN	840		
DB 781	SLFNGSHYLDVLYKMTADDQRYSGSTYLSDPRLTANGFKILIPGVSI TENYLEIEGMAN	840		
QY 841	CLPFYGVADLKEILNALNLRNAKEVYECRPRKVISYLEGEAVRLSRLQPLMYLSKEDIQDI	900		
DB 841	CLPFYGVADLKEILNALNLRNAKEVYECRPRKVISYLEGEAVRLSRLQPLMYLSKEDIQDI	900		
QY 901	IYRMKHQFCNEIKECVHGHPFFHLLTYLPETT	932		
DB 901	IYRMKHQFCNEIKECVHGHPFFHLLTYLPETT	932		

RESULT 2
 Q4VAL4 HUMAN
 ID Q4VAL4 HUMAN PRELIMINARY; PRT; 920 AA.
 AC Q4VAL4;

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 21, 2005, 20:03:01 ; Search time 47 Seconds
(without alignments)
1639.441 Million cell updates/sec

Title: US-10-079-429A-4
Perfect score: 4812
Sequence: 1 MKQLPAATVRLSSQIITS.....KECVHGRPFPHLTYLPETT 932

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 572060 seqs, 82675679 residues
Total number of hits satisfying chosen parameters: 572060

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA: *
1: /cgn2_6/ptodata/1/1aa/5 COMB pep.*
2: /cgn2_6/ptodata/1/1aa/6 COMB pep.*
3: /cgn2_6/ptodata/1/1aa/H COMB pep.*
4: /cgn2_6/ptodata/1/1aa/PTCUS COMB pep.*
5: /cgn2_6/ptodata/1/1aa/RE COMB pep.*
6: /cgn2_6/ptodata/1/1aa/baCfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4812	100.0	932	2	US-08-294-312B-4
2	4812	100.0	932	2	US-08-468-024B-4
3	4812	100.0	932	2	US-09-708-200-11
4	4812	100.0	932	2	US-08-465-679-4
5	4812	100.0	932	2	US-09-788-657-17
6	4812	100.0	932	2	US-09-788-657-18
7	4812	100.0	932	2	US-09-712-691-9
8	4812	100.0	932	2	US-10-641-068-17
9	4812	100.0	932	2	US-10-641-068-18
10	4805	99.9	932	2	US-09-707-468C-9
11	4693	97.5	921	2	US-09-949-016-7787
12	661	13.7	133	2	US-09-788-657-21
13	661	13.7	133	2	US-09-712-691-15
14	661	13.7	133	2	US-10-641-068-21
15	572	11.9	864	1	US-08-209-521-29
16	572	11.9	864	2	US-09-265-503B-138
17	569	11.8	856	2	US-08-709-784-2
18	569	11.8	862	1	US-08-209-521-23
19	569	11.8	862	1	US-08-209-521-30
20	569	11.8	862	2	US-09-059-461-2
21	569	11.8	862	2	US-08-961-810-133
22	569	11.8	862	2	US-08-352-902D-133
23	569	11.8	862	2	US-08-294-312B-6
24	569	11.8	862	2	US-08-468-024B-6
25	569	11.8	862	2	US-09-265-503B-133
26	569	11.8	862	2	US-09-708-200-9
27	569	11.8	862	2	US-08-465-679-6

28	569	11.8	862	2	US-09-712-691-7	Sequence 7, Appli
29	569	11.8	862	2	US-09-707-468C-7	Sequence 7, Appli
30	569	11.8	862	2	US-09-749-601A-11	Sequence 11, Appli
31	543.5	11.3	903	1	US-08-209-521-24	Sequence 24, Appl
32	543.5	11.3	903	2	US-08-961-810-134	Sequence 134, App
33	543.5	11.3	903	2	US-08-352-902D-134	Sequence 134, App
34	543.5	11.3	903	2	US-09-265-503B-134	Sequence 134, App
35	536	11.1	779	2	US-09-749-601A-12	Sequence 12, Appl
36	534.5	11.1	859	2	US-09-708-200-7	Sequence 7, Appli
37	534.5	11.1	859	2	US-09-788-657-16	Sequence 16, Appl
38	534.5	11.1	859	2	US-09-712-691-5	Sequence 5, Appli
39	534.5	11.1	859	2	US-09-707-468C-5	Sequence 5, Appli
40	534.5	11.1	859	2	US-10-641-068-16	Sequence 16, Appl
41	445.5	9.3	607	1	US-08-209-521-7	Sequence 7, Appli
42	445.5	9.3	607	2	US-08-961-810-3	Sequence 3, Appli
43	445.5	9.3	607	2	US-08-352-902D-3	Sequence 3, Appli
44	445.5	9.3	607	2	US-09-265-503B-3	Sequence 3, Appli
45	425.5	8.8	769	2	US-09-788-657-15	Sequence 15, Appl

ALIGNMENTS

RESULT 1
US-08-294-312B-4
; Sequence 4, Application US/08294312B
; Patent No. 6380369
; GENERAL INFORMATION:
; APPLICANT: Adams et al.
; TITLE OF INVENTION: Human DNA Mismatch Repair Proteins
; FILE REFERENCE: PFI06P2
; CURRENT APPLICATION NUMBER: US/08/294,312B
; CURRENT FILING DATE: 1994-08-23
; PRIOR APPLICATION NUMBER: 08/210,143
; PRIOR FILING DATE: 1994-03-16
; PRIOR APPLICATION NUMBER: 08/187,757
; PRIOR FILING DATE: 1994-01-27
; NUMBER OF SEQ ID NOS: 78
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4
; LENGTH: 932
; TYPE: PRT
; ORGANISM: homo sapiens
US-08-294-312B-4

Query Match	100.0%;	Score	4812;	DB 2;	Length	932;
Best Local Similarity	100.0%;	Pred. No.	0;			
Matches	932;	Conservative	0;	Mismatches	0;	Indels
QY	1	MKQLPAATVRLSSQIITSVVVVKELIENSLDAGATSDVVKLENYGFDKIEVRDNGEG	60			
Db	1	MKQLPAATVRLSSQIITSVVVVKELIENSLDAGATSDVVKLENYGFDKIEVRDNGEG	60			
QY	61	IKADAPVAMKYTTSKINSHEDLENLTYYGFRGALGSIICIAEVLITRITAADNFSTQ	120			
Db	61	IKADAPVAMKYTTSKINSHEDLENLTYYGFRGALGSIICIAEVLITRITAADNFSTQ	120			
QY	121	VYLDGSHILSQPHLGGTITVTLRLFKNLVPRKQFYSTAKCKDEIKKIODLLMSFG	180			
Db	121	VYLDGSHILSQPHLGGTITVTLRLFKNLVPRKQFYSTAKCKDEIKKIODLLMSFG	180			
QY	181	ILKPLDLRIVFVHNAKAVIWKSRVSDHKVMSVLGTAVNNMNSFOYHSEESQIYLSGFL	240			
Db	181	ILKPLDLRIVFVHNAKAVIWKSRVSDHKVMSVLGTAVNNMNSFOYHSEESQIYLSGFL	240			
QY	241	PKCDADHSFTSLSTPERSFIFINSRPVHQDKILKIRHYNLKLKESLRLVPVFLKID	300			
Db	241	PKCDADHSFTSLSTPERSFIFINSRPVHQDKILKIRHYNLKLKESLRLVPVFLKID	300			
QY	301	VPTADVNLTPDKSQVLLQNKESVLIALENLMTTCYGLPSTNSYNNKTDVSAADIVL	360			
Db	301	VPTADVNLTPDKSQVLLQNKESVLIALENLMTTCYGLPSTNSYNNKTDVSAADIVL	360			

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 21, 2005, 19:59:55 ; Search time 46 Seconds
(without alignments)
1949.435 Million cell updates/sec

Title: US-10-079-429A-4
Perfect score: 4812
Sequence: 1 MKQLPAATVRLSSSQIITS.....KECVHGRPFHHLTYLPETT 932

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 80:*
1: Pir1:*
2: Pir2:*
3: Pir3:*
4: Pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4812	100.0	932	2 S47597	mutL protein homol
2	587.5	12.2	805	2 T21957	hypothetical prote
3	569	11.8	862	2 S47598	mutL protein homol
4	556	11.6	904	2 S53896	DNA mismatch repai
5	536	11.1	779	2 T01304	hypothetical prote
6	529	11.0	794	2 T37989	DNA mismatch repai
7	448	9.3	669	2 C89904	DNA mismatch repai
8	424.5	8.8	769	2 S54525	mismatch repair pr
9	403	8.4	615	2 PH0853	methyl-directed mi
10	403	8.4	615	2 B91272	enzyme in methyl-d
11	403	8.4	615	2 B86113	enzyme in methyl-d
12	401.5	8.3	635	2 AC0046	DNA mismatch repai
13	396.5	8.2	516	2 H72427	DNA mismatch repai
14	394.5	8.2	629	2 E64046	mismatch repair pr
15	394	8.2	779	2 T25389	hypothetical prote
16	392	8.1	610	2 C70126	DNA mismatch repai
17	392	8.1	618	2 AG1048	DNA mismatch repai
18	392	8.1	618	2 A33588	mismatch repair pr
19	390.5	8.1	684	2 T50317	probable DNA misa
20	389.5	8.1	601	2 AD1250	DNA mismatch repai
21	387.5	8.1	603	2 AH1612	DNA mismatch repai
22	387.5	8.1	653	2 A82334	DNA mismatch repai
23	385	8.0	619	2 E82765	DNA mismatch repai
24	385	8.0	756	2 S43085	DNA mismatch repai
25	377.5	7.8	637	2 H83945	DNA mismatch repai
26	368	7.6	627	2 A69663	DNA mismatch repai
27	367	7.6	622	2 E97126	DNA mismatch repai
28	367	7.6	695	2 H86900	DNA mismatch repai
29	363.5	7.6	425	2 D70436	DNA mismatch repai

30	363.5	7.6	595	2 F71650	DNA mismatch repai
31	361.5	7.5	584	2 D84996	DNA mismatch repai
32	359.5	7.5	633	2 F83028	DNA mismatch repai
33	356	7.4	737	2 F85092	MHL protein [impo
34	356	7.4	737	2 T51620	DNA mismatch repai
35	355.5	7.4	576	2 A71497	probable DNA misa
36	355	7.4	649	2 H97891	DNA mismatch repai
37	355	7.4	649	2 A95020	DNA mismatch repai
38	352.5	7.3	658	2 C81860	DNA mismatch repai
39	350	7.3	649	2 A33589	mismatch repair pr
40	349.5	7.3	658	2 B81084	mismatch repair pr
41	344.5	7.2	580	2 A72032	DNA mismatch repai
42	344.5	7.2	580	2 B86592	DNA mismatch repai
43	344	7.1	576	2 G81657	hypothetical prote
44	340	7.1	695	2 S64862	DNA mismatch repai
45	332	6.9	610	2 A97870	DNA mismatch repai

ALIGNMENTS

RESULT 1

S47597
mutL protein homolog - human
C;Species: Homo sapiens (man)
C;Date: 27-Jan-1995 #sequence_revision 27-Jan-1995 #text_change 09-Jul-2004
C;Accession: S47597
R;Nicolaidis, N.C.; Papadopoulos, N.; Liu, B.; Wei, Y.F.; Carter, K.C.; Ruben, S.M.; R.
S.R.; Petersen, G.M.; de la Chapelle, A.; Vogelstein, B.; Kinzler, K.W.
Nature 371, 75-80, 1994
A;Title: Mutations of two PMS homologues in hereditary nonpolyposis colon cancer.
A;Reference number: S47597; MUID:94352394; PMID:8072530
A;Accession: S47597
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-932 <NIC>
A;Cross-references: UNIPROT:P54277; UNIPARC:UPI00000405F5; EMBL:U13695; NID:9535512; P
C;Genetics:
A;Gene: GDB:PMS1; PMSL1
A;Cross-references: GDB:386403; OMIM:600258
A;Map position: 2q31-2q33
F;571-643/Domain: HMG box homology <HMG1>

Query Match		100.0%;	Score 4812;	DB 2;	Length 932;
Best Local Similarity		100.0%;	Pred. No. 7.3e-231;		
Matches 932;		Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	MKQLPAATVRLSSSQIITSVSVVKELIENSLDAGATSVVDVKLENYGFDPKIEVRDNGEG	60		
Db	1	MKQLPAATVRLSSSQIITSVSVVKELIENSLDAGATSVVDVKLENYGFDPKIEVRDNGEG	60		
QY	61	IKAVDAPVAMKYITSKINSHEDLENLTYYGFRGEALGSCCIAEVLITTRTAADNPSTQ	120		
Db	61	IKAVDAPVAMKYITSKINSHEDLENLTYYGFRGEALGSCCIAEVLITTRTAADNPSTQ	120		
QY	121	YVLDSGSHLSQKPSHLGGTTVTLRLFKNLFPVRKQFYSTAKKCKDEIKKIQQLLMSFG	180		
Db	121	YVLDSGSHLSQKPSHLGGTTVTLRLFKNLFPVRKQFYSTAKKCKDEIKKIQQLLMSFG	180		
QY	181	ILKPDRLRIYFVNKAVIQKSRVSDHKMALMSVLGTAVMNMESFQVHSESQIYLSGFL	240		
Db	181	ILKPDRLRIYFVNKAVIQKSRVSDHKMALMSVLGTAVMNMESFQVHSESQIYLSGFL	240		
QY	241	PKCDADHSFTSLSTPERSIFINSRVHQKDILKLIRHHYNLCKLESTLYPVFFFLKID	300		
Db	241	PKCDADHSFTSLSTPERSIFINSRVHQKDILKLIRHHYNLCKLESTLYPVFFFLKID	300		
QY	301	VPTADVVDNLTPDKSOVLQNKESVLIALENLMTTCYGPLPSTNSYENKTDVSAADIVL	360		
Db	301	VPTADVVDNLTPDKSOVLQNKESVLIALENLMTTCYGPLPSTNSYENKTDVSAADIVL	360		
QY	361	SKTAETDVLFNKVSSESGKNYSNVDTSVIFPQNDMHNDESGKNTDCCINHQISTGDFGYGH	420		

Db 361 SIKTAETDVLFNKVESGKNSVNDTSVIFQNDMDNDESGKNTDCLNHQISIGDPGYCH 420
QY 421 CSSEISNIDKNTKNAFQDISMNSVWNSQTEYSKTCFISVVKHTQSENGKNDHIDESGE 480
Db 421 CSSEISNIDKNTKNAFQDISMNSVWNSQTEYSKTCFISVVKHTQSENGKNDHIDESGE 480
QY 481 NEEBAGLENSISADWESGKILKNSVGENIEPVKILVPEKSLPCKVNNNYPPEQWN 540
Db 481 NEEBAGLENSISADWESGKILKNSVGENIEPVKILVPEKSLPCKVNNNYPPEQWN 540
QY 541 LNEDESCNKSNIIDNKGKVTAYDLSNVRVKKPMSASALFYQDHRPQFLIENPKTSLED 600
Db 541 LNEDESCNKSNIIDNKGKVTAYDLSNVRVKKPMSASALFYQDHRPQFLIENPKTSLED 600
QY 601 ATLQTEELWKTISEBEKLYEKATKDLERYNSQMKRAIEQESOMSLKDGRRKIKPTSAW 660
Db 601 ATLQTEELWKTISEBEKLYEKATKDLERYNSQMKRAIEQESOMSLKDGRRKIKPTSAW 660
QY 661 NLAQKHKLKTSISNOPKDELQSQIEKRRSONIKMVQIPFSMNKLNKINFKKONKVDLEE 720
Db 661 NLAQKHKLKTSISNOPKDELQSQIEKRRSONIKMVQIPFSMNKLNKINFKKONKVDLEE 720
QY 721 KDEPCLIHNLRPDPDAWLTSTKTEVMLNPNRYVEEALLFKRLLENHKLPAEPELEKPTIMLPE 780
Db 721 KDEPCLIHNLRPDPDAWLTSTKTEVMLNPNRYVEEALLFKRLLENHKLPAEPELEKPTIMLPE 780
QY 781 SLFNGSHYDLVKMTADQRYSGSTYLSDPRLTANGFKIKLIPGVSIITENYLETEGMAN 840
Db 781 SLFNGSHYDLVKMTADQRYSGSTYLSDPRLTANGFKIKLIPGVSIITENYLETEGMAN 840
QY 841 CLPFFGVADLKELNALNNAKEVEYECPRKVISYLEGEAVRLSRQLPMYLSKEDIQDI 900
Db 841 CLPFFGVADLKELNALNNAKEVEYECPRKVISYLEGEAVRLSRQLPMYLSKEDIQDI 900
QY 901 IYRMKHQFQNEIKECVHGPRFFHHLTYLPETT 932
Db 901 IYRMKHQFQNEIKECVHGPRFFHHLTYLPETT 932
RESULT 2
T21957
hypothetical protein H12C20.2a - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C:Accession: T21957; T23069
R:Lennard, N.
submitted to the EMBL Data Library, June 1996
A:Reference number: Z19493
A:Accession: T21957
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-805 <WIL>
A:Cross-references: UNIPROT:Q9TVL8; UNIPARC:UPI0000081DBE; EMBL:Z74033; PIDN:CAA98478.1
A:Experimental source: clone F38B7
R:White, S.
submitted to the EMBL Data Library, March 1998
A:Reference number: Z19670
A:Accession: T23069
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-805 <W12>
A:Cross-references: UNIPARC:UPI0000081DBE; EMBL:AL022272; PIDN:CAAL8355.1; GSPDB:GN00023
A:Experimental source: clone H12C20
C:Genetics:
A:Gene: CBSP.H12C20.2a
A:Map position: 5
A:Introns: 21/3; 109/2; 267/2; 300/3; 329/1; 393/2; 553/3; 612/2; 668/2; 724/1; 783/3
C:Superfamily: DNA mismatch repair protein

Query Match 12.2%; Score 587.5; DB 2; Length 805;
Best Local Similarity 22.9%; Pred. No. 1.1e-21;
Matches 226; Conservative 164; Mismatches 326; Indels 271; Gaps 37;

RESULT 3

S47598

mutL protein homolog - human

C:Species: Homo sapiens (man)

QY 12 LSSQIITSVSVVVKELIENSLDAGATSDVKLENYGFQKIEVRDNGEGIKAVDAPVMAM 71
Db 17 LTTAQVVSLSAIRQLDINSIDAGSTIIDIRVKNKGESIEVQDNGSGIEARNFDALC 76
QY 72 KYTTSKINSHEDLENLTYYGRGEALGSCCIAEVLITRRTAADNFSQYVLDGSGHILS 131
Db 77 PHSTSKLTQFSDFKLATLGRGEALNACTVSSVSIPTASDTEIGTALTYDHSIGNITC 136
QY 132 QKPSHLGGTWTALRFLNLPVRKQFYSTAKCKDEIKIQDILLMSFGILKPDRLRIVFV 191
Db 137 QSAARELGTTIIVNKLFTLPPVRK--ELERSQREFVKKLLSTVQSFALLCPHKLICT 194
QY 192 HNKAVIQKSRVSDHOMALMSVLG--TAVMN----- 220
Db 195 NN-----INGKTNLTCTPGGTTISIQDVVANLFGIARKIENSKIGSLIPIQONQP 245
QY 221 NMESQYHS---ESQIY----LSGFLPKCADHSFTSLSTPERSFIFINSRPVHQKDL 273
Db 246 DVEIMTHSVPMEEHFFDLFKIRGVSSC--EHG-CGRGTSDDRFVYINNRPVYSRVC 302
QY 274 KLIRHHYLNKCLKESTRILYPVFFLKIDVPTADVDVNLTPDKSQVLLQNKESVLIALENLM 333
Db 303 SVINDVYK---QFNKKQYPIIVLFDVPPPEKIDVNVTPDKKTVMLERHLLAVVRASM 358
QY 334 TTCYGLPSTNSYENKKTVDVSAADIV--LSKTAETDVLFNKVESGKNSVNDTSVIPPON 392
Db 359 MKTY--LAKIVGSHSVTRSSVEDRRIMNLSQQSFNSASFMSSKSTPDDPNNTT-----LN 411
QY 393 DMHNDSEKNTDDCLNHQISIGDFYGHCSSEISNIDKNTKNAFQDISNSVNSWNSQTE 452
Db 412 STYPEDSLNLTSDLL-----KORK-----ENRSP 436
QY 453 YSKTCFISVVKHTQSENGKNDHIDEGSENEEAGLENS--EIS----- 494
Db 437 AKKSC--PMIRTE-----PHSVSTSNRSRTORLENFSTMEPKRVEVSKKIPSKDKK 489
QY 495 -ADEWSRGNILKNS-----VGENIEPVKILVPKSLPCKVNNNYPPEQMINNESCNKK 549
Db 490 LTDEELRSVATEENPLKXAGE-IDDIEL--EQS-----QESQDVNESQCSQD 534
QY 550 SNVIDNKGKVTAYDLSNVRVKKPMSASALFYQDHRPQFLIENPKTSLEDATLQIEELW 609
Db 535 SQTSON--SRVSFTLRPQKIKFSMK--LLREAYSPK--TDETDNDTEEAESVSAE--- 584
QY 610 KTLSEEEKLYEKATKDLERYNSQMKRAIEQESQMSLKDGRKKIKPTSAWLAQKHKL 669
Db 585 KDVLNEITTKINKEENDDAER---QLSRSL-----TKDDFSKMKLIQGFN----- 626
QY 670 TSLSNQPKLDELLOSQIEKRRSQNIKMVQIPFSMNKLNKINFKKONKVDLEEKDEPCLIH 729
Db 627 -----HGFIICR 633
QY 730 LRFPDAWLTSTKTEVMLNPNRYVEEALLFKRLLENHKLPAEPELEKPTIMLTESIENGSHVL 789
Db 634 LR-----GHLFIVDQHASDEKYNFERLOQSSAKLTKQFLFMPTALG----FGAVQEL 680
QY 790 DVLVYKMTADQRYSGSTYLSDPRLTANGFKIKLIPGVSIITENYLEI-----EGMANCLP 843
Db 681 IIRENL-----PIFANGDFPFESENDGCIKTFLTARPPELLAQCLTN--- 722
QY 844 FYGVADLKELNALNNAKEVEYECPRKVISYLEGEAVRLSRQLPMYLSKEDIQDIIVR 903
Db 723 -----SDLEEL--AVVSQVPMQWY--RPVRIKIFASKACKRKSVMIGKPLNQREMTQIIRH 775
QY 904 MKHQFGNEIKECVHGPRFFHHLTYLPE 930
Db 776 LAKL--DQPMWNCPHGRPTIRHLASLFD 800

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 21, 2005, 19:58:15 ; Search time 190 seconds

(without alignments)
2155.269 Million cell updates/sec

Title: US-10-079-429A-4

Perfect score: 4812

Sequence: 1 MKQLPAATVRLSSSQIITS.....KECVHGRPFPHLTLPETT 932

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2443163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters: 2443163

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq 21:*

- 1: geneseqp1980s:*
- 2: geneseqp1990s:*
- 3: geneseqp2000s:*
- 4: geneseqp2001s:*
- 5: geneseqp2002s:*
- 6: geneseqp2003as:*
- 7: geneseqp2003bs:*
- 8: geneseqp2004s:*
- 9: geneseqp2005s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4812	100.0	932	4	AAB85852 Human PMS
2	4812	100.0	932	4	AAG63953 Amino aci
3	4812	100.0	932	4	AAG63954 Amino aci
4	4812	100.0	932	5	AAG98776 Human pos
5	4812	100.0	932	5	AAG24357 Human mis
6	4812	100.0	932	5	AAG18553 Human mis
7	4812	100.0	932	5	AAG28277 Human MLH
8	4812	100.0	932	5	AAG24684 Human PMS
9	4812	100.0	932	6	ABU07972 Human PMS
10	4812	100.0	932	6	ABU07971 Human PMS
11	4812	100.0	932	6	ABU89659 Human PMS
12	4812	100.0	932	6	ABU89660 Human PMS
13	4812	100.0	932	6	ABO07415 Human Mut
14	4812	100.0	932	6	AAO27514 Human mis
15	4812	100.0	932	6	AAO27515 Human mis
16	4812	100.0	932	7	ADA06246 Human mis
17	4812	100.0	932	7	ADA06244 Human mis
18	4812	100.0	932	7	ADC89605 Human PMS
19	4812	100.0	932	7	ADC89607 Human PMS
20	4812	100.0	932	7	ADFI7892 Human PMS
21	4812	100.0	932	7	ADG62881 Human PMS
22	4812	100.0	932	7	ADG62882 Human PMS
23	4812	100.0	932	7	ADH62629 Human mis
24	4812	100.0	932	7	ADH60983 Human mis

ALIGNMENTS

RESULT 1

AAB85852
ID AAB85852 standard; protein; 932 AA.

XX AC AAB85852;

XX DT 29-OCT-2001 (first entry)

XX DE Human PMS1 protein.

XX KW Hypermutable bacteria; mismatch repair gene; MMR gene; MutH; MutS; MutL;
XX KW MutY; PMS2; MLH1; MLH3; PMSR; biocatalysis; bioremediation; biochemical;
XX KW drug discovery; detoxification; toxin; biotransformation; PMS1.

XX OS Homo sapiens.

XX PN WO200159092-A2.

XX PD 16-AUG-2001.

XX PF 12-FEB-2001; 2001WO-US004339.

XX PR 11-FEB-2000; 2000US-0181929P.

XX PA (UYJO) UNIV JOHNS HOPKINS.

XX PI Nicolaides NC, Sass PM, Grasso L, Vogelstein B, Kinzler KW;

XX WPI; 2001-514664/56.

XX N-PSDB; AAH76365.

XX PT Making hypermutable bacteria for biocatalysis, bioremediation and drug
XX PT discovery, involves introducing polynucleotide comprising dominant
XX PT negative allele of mismatch repair gene under regulatory sequence
XX PT control.

XX PS Example 1; Page 41; 68pp; English.

XX CC The invention provides a method for generating a hypermutable bacteria.
XX CC The method involves introducing a polynucleotide having a dominant
XX CC negative allele of a mismatch repair (MMR) gene under the control of an
XX CC inducible transcription regulatory sequence, into a bacterium. The cell
XX CC becomes inducibly hypermutable. The method is useful to create desirable
XX CC output traits for commercial applications, using dominant negative
XX CC alleles of mismatch repair proteins. The mismatch repair gene is a MutH,
XX CC MutS, MutL or MutY homologue and can be selected from PMS2, MLH1, MLH3,
XX CC PMSR or PMSR homologue. The hypermutable bacteria is useful for the

25 4812 100.0 932 7 ADJ68675
26 4812 100.0 932 8 ADF78840 Human hea
27 4812 100.0 932 8 ADG46767 Human MMR
28 4812 100.0 932 8 ADO40066 Human PMS
29 4812 100.0 932 8 ADP66682 Human mis
30 4812 100.0 932 8 ADP66680 Human mis
31 4812 100.0 932 8 ADL13883 Human DNA
32 4812 100.0 932 8 ADU77038 Human mis
33 4812 100.0 932 9 ADX08123 Cyclin-de
34 4812 100.0 932 9 ADX58473 Amino aci
35 4812 100.0 932 9 ADY53431 Human PMS
36 4807 99.9 931 2 AAR79009 Human DNA
37 4805 99.9 932 8 ADT98686 Human PMS
38 4805 99.9 932 8 ADU77022 Human mis
39 4627 96.2 900 4 AAB85851 Human PMS
40 4317.5 89.7 847 8 ABM83683 Human dia
41 1168 24.3 232 5 AAE28280 Human MLH
42 661 13.7 133 4 AAB85855 Human PMS
43 661 13.7 133 4 AAG63957 Amino aci
44 661 13.7 133 5 AAO18559 Human tru
45 661 13.7 133 5 AAE24687 Human PMS

CC production, biocatalysis, bioremediation and drug discovery. It is also
CC useful in manufacturing industry for the generation of new biochemicals
CC useful for detoxifying noxious chemicals from by-products of
CC manufacturing processes or those used as catalysts, for remediation of
CC toxins present in the environment including polychlorobenzenes, heavy
CC metals and other environmental hazards for which there is a need to
CC remove them from the environment. The hypermutable bacteria is also
CC useful for screening novel mutations in a gene or a set of genes that
CC produce variant siblings that exhibit a new output trait not found in
CC wild type cells. The bacteria are also useful for producing increased
CC quantity or quality of protein or non-protein therapeutic molecule e.g.
CC Penicillin G, Erythromycin and Clavulanic acid, by biotransformation.
CC Dominant negative alleles of the MMR gene are useful for producing higher
CC quantities of recombinant polypeptides. The present sequence represents a
CC human PMS1 protein
XX
SQ Sequence 932 AA;

Query Match 100.0%; Score 4812; DB 4; Length 932;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 932; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKQLPAATVRLSSQIITSVSVVKELIENSIDAGATSDVKLENYGDKIEVRDNGSG 60
DB 1 MKQLPAATVRLSSQIITSVSVVKELIENSIDAGATSDVKLENYGDKIEVRDNGSG 60

QY 61 IKAVDAPVAMKYTSKINSHEDLENLTYYGPRGEALGSGICIAEVLITRTRAADNFSTQ 120
DB 61 IKAVDAPVAMKYTSKINSHEDLENLTYYGPRGEALGSGICIAEVLITRTRAADNFSTQ 120

QY 121 YLDSGSHLTSQKPHLGGTGTVALRFLPNLPVRKQFYSTAKCKDEIKKIQDILLMSFG 180
DB 121 YLDSGSHLTSQKPHLGGTGTVALRFLPNLPVRKQFYSTAKCKDEIKKIQDILLMSFG 180

QY 181 ILKPDRLRVFNKAVIQKSRVSDHKMALMSVLGTAVNNMESFYHSEESQIYLSGFL 240
DB 181 ILKPDRLRVFNKAVIQKSRVSDHKMALMSVLGTAVNNMESFYHSEESQIYLSGFL 240

QY 241 PKCDADHSPTSLSSTPERSFIFNSRPVHOKDILKLRHHYNLKLKESTRLPVFFFLKID 300
DB 241 PKCDADHSPTSLSSTPERSFIFNSRPVHOKDILKLRHHYNLKLKESTRLPVFFFLKID 300

QY 301 VPTADVNLTPDKSOVLQNKESVLIALENLMTTCYGPLPSTNSYNNKTDVSAADIVL 360
DB 301 VPTADVNLTPDKSOVLQNKESVLIALENLMTTCYGPLPSTNSYNNKTDVSAADIVL 360

QY 361 SKTAETDVLFNKVESGKNYSNVDTSVIPFQNDMHNDESGKNTDCLNHQISIGDFGYGH 420
DB 361 SKTAETDVLFNKVESGKNYSNVDTSVIPFQNDMHNDESGKNTDCLNHQISIGDFGYGH 420

QY 421 CSSEISNIDKNTKNAFQDISMSNVSWNSQTEYSKTCFISSVKHTQSENGKNDHIDESGE 480
DB 421 CSSEISNIDKNTKNAFQDISMSNVSWNSQTEYSKTCFISSVKHTQSENGKNDHIDESGE 480

QY 481 NEEBAGLENSSEISADEWGRNLIKNSVGENTEPVKILVPEKSLPCKVSNNNYPPEQNN 540
DB 481 NEEBAGLENSSEISADEWGRNLIKNSVGENTEPVKILVPEKSLPCKVSNNNYPPEQNN 540

QY 541 LNEДСNKKSNNIDNKGSKVTAYDILLSNRVVKPMASASALFVQDHRPQFLIENPKTSLSD 600
DB 541 LNEДСNKKSNNIDNKGSKVTAYDILLSNRVVKPMASASALFVQDHRPQFLIENPKTSLSD 600

QY 601 ATLQIEELWKTLSSEBKLYEEKATKDLERYNSQMKRAIEQESQMSLKDGRKKIKPTSAW 660
DB 601 ATLQIEELWKTLSSEBKLYEEKATKDLERYNSQMKRAIEQESQMSLKDGRKKIKPTSAW 660

QY 661 NLAQKHKLTSLSNQPKLDELLOSQIEKRSQNKVQIPFSMKNLKNFKQNKVDLEE 720
DB 661 NLAQKHKLTSLSNQPKLDELLOSQIEKRSQNKVQIPFSMKNLKNFKQNKVDLEE 720

QY 721 KDEPCLIHNLRPDPAWMTSKTEVMLNLPYRVEEALLFKRLLENHKLPAEPLKPTMLTE 780
DB 721 KDEPCLIHNLRPDPAWMTSKTEVMLNLPYRVEEALLFKRLLENHKLPAEPLKPTMLTE 780

QY 781 SLFNGSHYLDVLYKMTADDQRYSGSTYLSDRLTANGFKKILPGVSITENYLEIEGMAN 840
DB 781 SLFNGSHYLDVLYKMTADDQRYSGSTYLSDRLTANGFKKILPGVSITENYLEIEGMAN 840

QY 841 CLPFYGVADLKEILNALNRNAKEVYECRPRKVISYLEGEAVRLSRQLPMYLSKEDIQDI 900
DB 841 CLPFYGVADLKEILNALNRNAKEVYECRPRKVISYLEGEAVRLSRQLPMYLSKEDIQDI 900

QY 901 IYRMKHQFGNEIKECVHGRPFPHLTLYLPETT 932
DB 901 IYRMKHQFGNEIKECVHGRPFPHLTLYLPETT 932

RESULT 2
AAG63953 standard; protein; 932 AA.
XX
AC AAG63953;
XX
DT 29-OCT-2001 (first entry)
XX
DE Amino acid sequence of human mismatch repair protein PMS2.
XX
KW PMS2; mismatch repair gene; MMR gene; hypermutable yeast.
XX
OS Homo sapiens.
XX
FN WO200162945-A1.
XX
PD 30-AUG-2001.
XX
PF 21-FEB-2001; 2001WO-US005447.
XX
PR 23-FEB-2000; 2000US-0184336P.
XX
PA (UYJO) UNIV JOHNS HOPKINS.
PA (NICO/) NICOLAIDES N C.
PA (SASS/) SASS P M.
PA (GRAS/) GRASSO L.
PA (VOGE/) VOGELSTEIN B.
PA (KINZ/) KINZLER K W.
XX
PI Nicolaides NC, Sass PM, Grasso L, Vogelstein B, Kinzler KW;
XX
XX WPI; 2001-522820/57.
XX
DR N-PSDB; AAH75041.
XX
PT Making hypermutable yeast that exhibit novel selected output traits for
PT commercial applications, comprises introducing polynucleotide containing
PT dominant negative allele of mismatch repair gene.
PS
PS Disclosure; Page 38; 60pp; English.
CC
CC The present sequence represents human PSM2. PMS2 is a mismatch repair
CC (MMR) gene. The specification describes a method for making a
CC hypermutable yeast, comprising introducing a polynucleotide containing a
CC dominant negative allele of a mismatch repair (MMR) gene, into a yeast,
CC whereby the cell becomes hypermutable. The method is useful to create
CC desirable output traits for commercial applications, using dominant
CC negative alleles of mismatch repair proteins. The hypermutable yeast is
CC useful for production, biocatalysis, bioremediation and drug discovery.
CC It is also useful in genetic screens for the direct selection of variant
CC subclones that exhibit new output traits. The hypermutable yeast is also
CC useful in the manufacturing industry for the generation of new
CC biochemicals, for detoxifying noxious chemicals from by-products of
CC manufacturing processes or those used as catalysts, for remediation of
CC toxins present in the environment including polychlorobenzenes, heavy
CC metals and other environmental hazards for which there is a need to
CC remove them from the environment. The yeast is also useful for producing
CC increased quantity or quality of protein or non-protein therapeutic
CC molecule e.g., Penicillin G, Erythromycin and Clavulanic acid, by
CC biotransformation

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 21, 2005, 20:06:26 ; Search time 167 seconds
(without alignments)
2331.838 Million cell updates/sec

Title: US-10-079-429A-4
Perfect score: 4812
Sequence: 1 MKQLPAATVRLSSSQIITS.....KECVHGRPFPHHLYLPETT 932

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 1867569

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : Published Applications AA Main:
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pcp.*
2: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pcp.*
3: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pcp.*
4: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pcp.*
5: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pcp.*
6: /cgn2_6/ptodata/1/pubpaa/US11_PUBCOMB.pcp.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4812	100.0	932	3	US-09-788-657-17
2	4812	100.0	932	3	US-09-788-657-18
3	4812	100.0	932	3	US-09-912-697-6
4	4812	100.0	932	3	US-09-912-697-8
5	4812	100.0	932	3	US-09-760-285-18
6	4812	100.0	932	4	US-10-079-429-4
7	4812	100.0	932	4	US-10-270-839-27
8	4812	100.0	932	4	US-10-270-839-29
9	4812	100.0	932	4	US-10-243-130-9
10	4812	100.0	932	4	US-10-371-857-3
11	4812	100.0	932	4	US-10-371-857-4
12	4812	100.0	932	4	US-10-371-634-5
13	4812	100.0	932	4	US-10-371-634-7
14	4812	100.0	932	4	US-10-348-074-3
15	4812	100.0	932	4	US-10-348-074-5
16	4812	100.0	932	4	US-10-369-845-11
17	4812	100.0	932	4	US-10-641-068-17
18	4812	100.0	932	4	US-10-641-068-18
19	4812	100.0	932	4	US-10-408-765A-481
20	4812	100.0	932	4	US-10-813-502-9
21	4812	100.0	932	4	US-10-714-228-2
22	4812	100.0	932	4	US-10-714-228-4
23	4812	100.0	932	5	US-10-933-034-2
24	4812	100.0	932	5	US-10-901-650-9
25	4812	100.0	932	5	US-11-056-776-13
26	4805	99.9	932	5	US-10-850-370-9
27	661	13.7	133	3	US-09-788-657-21

28	661	13.7	133	3	US-09-912-697-14	Sequence 14, Appl
29	661	13.7	133	3	US-09-760-285-24	Sequence 24, Appl
30	661	13.7	133	4	US-10-270-839-35	Sequence 35, Appl
31	661	13.7	133	4	US-10-371-857-12	Sequence 12, Appl
32	661	13.7	133	4	US-10-371-634-13	Sequence 13, Appl
33	661	13.7	133	4	US-10-348-074-11	Sequence 11, Appl
34	661	13.7	133	4	US-10-641-068-21	Sequence 21, Appl
35	661	13.7	133	4	US-10-714-228-6	Sequence 6, Appl
36	615.5	12.8	899	6	US-11-097-143-15705	Sequence 15705, A
37	607	12.6	923	4	US-10-270-839-45	Sequence 45, Appl
38	607	12.6	923	4	US-10-714-228-36	Sequence 36, Appl
39	607	12.6	923	5	US-10-933-034-36	Sequence 36, Appl
40	607	12.6	923	5	US-10-933-034-36	Sequence 36, Appl
41	587.5	12.2	805	4	US-10-369-493-6498	Sequence 6498, Ap
42	587.5	12.2	805	4	US-10-369-493-6498	Sequence 6498, Ap
43	585.5	12.2	1013	4	US-10-425-114-62661	Sequence 62661, A
44	582.5	12.1	990	4	US-10-425-115-362659	Sequence 362659,
45	572	11.9	864	4	US-10-349-607-138	Sequence 138, App

ALIGNMENTS

RESULT 1
US-09-788-657-17
; Sequence 17, Application US/09788657
; Patent No. US20020123149A1
; GENERAL INFORMATION:
; APPLICANT: Nicolaides, Nicholas
; APPLICANT: Sasse, Philip
; APPLICANT: Kinzler, Kenneth
; APPLICANT: Grasso, Luigi
; APPLICANT: Vogelstein, Bert
; TITLE OF INVENTION: Methods for generating hypermutable
; TITLE OF INVENTION: yeast
; FILE REFERENCE: 01107.00097
; CURRENT APPLICATION NUMBER: US/09/788,657
; CURRENT FILING DATE: 2001-02-21
; PRIOR APPLICATION NUMBER: 60/184,336
; PRIOR FILING DATE: 2000-02-23
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 17
; LENGTH: 932
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-788-657-17

Query Match	100.0%	Score 4812;	DB 3;	Length 932;
Best Local Similarity	100.0%	Pred. No. 1.5e-285;		
Matches	932;	Conservative	0;	Mismatches 0;
Indels	0;	Gaps	0;	
Qy	1	MKQLPAATVRLSSSQIITSVVSVVKELIENSLDAGATSDVKLENYGFDKIEVRDNGEG	60	
Db	1	MKQLPAATVRLSSSQIITSVVSVVKELIENSLDAGATSDVKLENYGFDKIEVRDNGEG	60	
Qy	61	IKAVDAPVWAMKYTSKINSHEDLENTTYGFRGALGSIICIAEVLITTTAAADNFSTQ	120	
Db	61	IKAVDAPVWAMKYTSKINSHEDLENTTYGFRGALGSIICIAEVLITTTAAADNFSTQ	120	
Qy	121	YVLDSGHILSQKPSHLGGQTTVTLRLFKNLVPRKQFYSTAKCKDKIKKIQDILLMSFG	180	
Db	121	YVLDSGHILSQKPSHLGGQTTVTLRLFKNLVPRKQFYSTAKCKDKIKKIQDILLMSFG	180	
Qy	181	ILKPDRLRIVFVHNKAVIWKSRVSDHKMALMSVLGTAVMNNMESFYHSESIYLSGFL	240	
Db	181	ILKPDRLRIVFVHNKAVIWKSRVSDHKMALMSVLGTAVMNNMESFYHSESIYLSGFL	240	
Qy	241	PKCDADHSFTSLSTPERSFIINSRPVHQDKILKIRHHYNLKCILKESTRLYPVFFLKID	300	
Db	241	PKCDADHSFTSLSTPERSFIINSRPVHQDKILKIRHHYNLKCILKESTRLYPVFFLKID	300	
Qy	301	VPTADVVDNLTPDKSQVLLQNKESVLIENLMTTCYGLPSTNSYENNKTDVSAADIVL	360	

Db 301 VPTADVNLTPDKSQVLLQNKESVLIJALENLMTTCYGLPSTNSYNNKTDVSAADIVL 360
Qy 361 SKTAETDVLNFKVSSGKNYSNVDTSVIPFQNDMHNDESGKNTDCLNHQISIGDFGYGH 420
Db 361 SKTAETDVLNFKVSSGKNYSNVDTSVIPFQNDMHNDESGKNTDCLNHQISIGDFGYGH 420
Qy 421 CSSEISNIDKNTKNAFQDISMSNVSWNSQTEYSKTCFISVVKHTQSENGKNDHIDEGE 480
Db 421 CSSEISNIDKNTKNAFQDISMSNVSWNSQTEYSKTCFISVVKHTQSENGKNDHIDEGE 480
Qy 481 NEEEAGLENSSEISADEWSRGNILKNSVGENIEPVKILVPEKSLPCKVSNNNYPPEQWN 540
Db 481 NEEEAGLENSSEISADEWSRGNILKNSVGENIEPVKILVPEKSLPCKVSNNNYPPEQWN 540
Qy 541 LNEDESCNKSNNVDNKSQVLTAYDLSNRVVIKPKMSASALFVQDHRPQPLIENPKTSLD 600
Db 541 LNEDESCNKSNNVDNKSQVLTAYDLSNRVVIKPKMSASALFVQDHRPQPLIENPKTSLD 600
Qy 601 ATLQIEELWKTLSBEEKLYBEEKATKDLERYNSQMKRAIEQBSQMSLKDGRKKIKPTSAW 660
Db 601 ATLQIEELWKTLSBEEKLYBEEKATKDLERYNSQMKRAIEQBSQMSLKDGRKKIKPTSAW 660
Qy 661 NLAQKHKLKTSLSNQPKLDELLOQSIKERRSQNIKMVQIPFSMKNLKINFKKQNKVDLEE 720
Db 661 NLAQKHKLKTSLSNQPKLDELLOQSIKERRSQNIKMVQIPFSMKNLKINFKKQNKVDLEE 720
Qy 721 KDEPCLIHNLRFPPDAWLMTSKTEVMLLNPRYVEEALLFKRLLENHKLPAEPLKPIMLTE 780
Db 721 KDEPCLIHNLRFPPDAWLMTSKTEVMLLNPRYVEEALLFKRLLENHKLPAEPLKPIMLTE 780
Qy 781 SLFNGSHYLDVLYKMTADDQRYSGSTYLSDPRLTANGFKIKLIPGVSIITENYLETEGMAN 840
Db 781 SLFNGSHYLDVLYKMTADDQRYSGSTYLSDPRLTANGFKIKLIPGVSIITENYLETEGMAN 840
Qy 841 CLPFFGVADLKEILNAILNRNAKEVYECRPRKVIISYLEGEAVRLSRQLPMLSKEDIQDI 900
Db 841 CLPFFGVADLKEILNAILNRNAKEVYECRPRKVIISYLEGEAVRLSRQLPMLSKEDIQDI 900
Qy 901 IYRMKHQFNGEIKECVHGRRPFFHHLTYLPETT 932
Db 901 IYRMKHQFNGEIKECVHGRRPFFHHLTYLPETT 932

RESULT 2

US-09-788-657-18
; Sequence 18, Application US/09788657
; Patent No. US20020123149A1
; GENERAL INFORMATION:
; APPLICANT: Nicolaides, Nicholas
; APPLICANT: Sasse, Philip
; APPLICANT: Kinzler, Kenneth
; APPLICANT: Grasso, Luigi
; APPLICANT: Vogelstein, Bert
; TITLE OF INVENTION: Methods for generating hypermutable
; FILE REFERENCE: 01107.00097
; CURRENT APPLICATION NUMBER: US/09/788,657
; CURRENT FILING DATE: 2001-02-21
; PRIOR APPLICATION NUMBER: 60/184,336
; PRIOR FILING DATE: 2000-02-23
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 18
; LENGTH: 932
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-788-657-18

Query Match 100.0%; Score 4812; DB 3; Length 932;
Best Local Similarity 100.0%; Pred. No. 1.5e-285;
Matches 932; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MKQLPAATVRLSSQIITSVVVKELIENSLDAGATSDVKLENYGFDKIEVRDNGEG 60
Db 1 MKQLPAATVRLSSQIITSVVVKELIENSLDAGATSDVKLENYGFDKIEVRDNGEG 60
Qy 61 IKAVDAPVWAMKYTTSKINSHEDLENLTYYFGRGALGSI CCIAEVLITTRTAADNFSTQ 120
Db 61 IKAVDAPVWAMKYTTSKINSHEDLENLTYYFGRGALGSI CCIAEVLITTRTAADNFSTQ 120
Qy 121 YVLDSGSHILSOKPSHILGOGTTVLTALRFLKPLVVRKQFYSTAKCKDIKKIQDILLMSFG 180
Db 121 YVLDSGSHILSOKPSHILGOGTTVLTALRFLKPLVVRKQFYSTAKCKDIKKIQDILLMSFG 180
Qy 181 ILKPDRLRVFVHNKAVIWKSRVSDHKWALMSVLGTA VMNNMESQYHSEESQIYLSGFL 240
Db 181 ILKPDRLRVFVHNKAVIWKSRVSDHKWALMSVLGTA VMNNMESQYHSEESQIYLSGFL 240
Qy 241 PKCDADHSFTSLSTPERSFIFINSRPVHQDKILKIRHHYMLKCLKESTRLYPVFPLKID 300
Db 241 PKCDADHSFTSLSTPERSFIFINSRPVHQDKILKIRHHYMLKCLKESTRLYPVFPLKID 300
Qy 301 VPTADVNLTPDKSQVLLQNKESVLIJALENLMTTCYGLPSTNSYNNKTDVSAADIVL 360
Db 301 VPTADVNLTPDKSQVLLQNKESVLIJALENLMTTCYGLPSTNSYNNKTDVSAADIVL 360
Qy 361 SKTAETDVLNFKVSSGKNYSNVDTSVIPFQNDMHNDESGKNTDCLNHQISIGDFGYGH 420
Db 361 SKTAETDVLNFKVSSGKNYSNVDTSVIPFQNDMHNDESGKNTDCLNHQISIGDFGYGH 420
Qy 421 CSSEISNIDKNTKNAFQDISMSNVSWNSQTEYSKTCFISVVKHTQSENGKNDHIDEGE 480
Db 421 CSSEISNIDKNTKNAFQDISMSNVSWNSQTEYSKTCFISVVKHTQSENGKNDHIDEGE 480
Qy 481 NEEEAGLENSSEISADEWSRGNILKNSVGENIEPVKILVPEKSLPCKVSNNNYPPEQWN 540
Db 481 NEEEAGLENSSEISADEWSRGNILKNSVGENIEPVKILVPEKSLPCKVSNNNYPPEQWN 540
Qy 541 LNEDESCNKSNNVDNKSQVLTAYDLSNRVVIKPKMSASALFVQDHRPQPLIENPKTSLD 600
Db 541 LNEDESCNKSNNVDNKSQVLTAYDLSNRVVIKPKMSASALFVQDHRPQPLIENPKTSLD 600
Qy 601 ATLQIEELWKTLSBEEKLYBEEKATKDLERYNSQMKRAIEQBSQMSLKDGRKKIKPTSAW 660
Db 601 ATLQIEELWKTLSBEEKLYBEEKATKDLERYNSQMKRAIEQBSQMSLKDGRKKIKPTSAW 660
Qy 661 NLAQKHKLKTSLSNQPKLDELLOQSIKERRSQNIKMVQIPFSMKNLKINFKKQNKVDLEE 720
Db 661 NLAQKHKLKTSLSNQPKLDELLOQSIKERRSQNIKMVQIPFSMKNLKINFKKQNKVDLEE 720
Qy 721 KDEPCLIHNLRFPPDAWLMTSKTEVMLLNPRYVEEALLFKRLLENHKLPAEPLKPIMLTE 780
Db 721 KDEPCLIHNLRFPPDAWLMTSKTEVMLLNPRYVEEALLFKRLLENHKLPAEPLKPIMLTE 780
Qy 781 SLFNGSHYLDVLYKMTADDQRYSGSTYLSDPRLTANGFKIKLIPGVSIITENYLETEGMAN 840
Db 781 SLFNGSHYLDVLYKMTADDQRYSGSTYLSDPRLTANGFKIKLIPGVSIITENYLETEGMAN 840
Qy 841 CLPFFGVADLKEILNAILNRNAKEVYECRPRKVIISYLEGEAVRLSRQLPMLSKEDIQDI 900
Db 841 CLPFFGVADLKEILNAILNRNAKEVYECRPRKVIISYLEGEAVRLSRQLPMLSKEDIQDI 900
Qy 901 IYRMKHQFNGEIKECVHGRRPFFHHLTYLPETT 932
Db 901 IYRMKHQFNGEIKECVHGRRPFFHHLTYLPETT 932

RESULT 3

US-09-912-697-6
; Sequence 6, Application US/09912697
; Publication No. US20030068808A1
; GENERAL INFORMATION:
; APPLICANT: Nicolaides, Nicholas C
; APPLICANT: Sasse, Philip M
; APPLICANT: Grasso, Luigi M

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 21, 2005, 20:11:22 ; Search time 13 Seconds
(without alignments)
511.317 Million cell updates/sec

Title: US-10-079-429A-4
Perfect score: 4812
Sequence: 1 MKQLPAATVRLSSSQIITS.....KECVHGRPFPHHLYLPTT 932

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 53982 seqs, 7132107 residues

Total number of hits satisfying chosen parameters: 53982

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA New:
1: /cgn2_6/ptodata/2/pubpaa/US08 NEW PUB.pap.*
2: /cgn2_6/ptodata/2/pubpaa/US06 NEW PUB.pap.*
3: /cgn2_6/ptodata/2/pubpaa/US07 NEW PUB.pap.*
4: /cgn2_6/ptodata/2/pubpaa/PCT_NEW PUB.pap.*
5: /cgn2_6/ptodata/2/pubpaa/US09 NEW PUB.pap.*
6: /cgn2_6/ptodata/2/pubpaa/US10 NEW PUB.pap.*
7: /cgn2_6/ptodata/2/pubpaa/US11 NEW PUB.pap.*
8: /cgn2_6/ptodata/2/pubpaa/US60 NEW PUB.pap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4812	100.0	932	7	US-11-188-743-17
2	4812	100.0	932	7	US-11-188-743-18
3	661	13.7	133	7	US-11-188-743-21
4	569	11.8	862	7	US-11-128-420-11
5	536	11.1	779	7	US-11-128-420-12
6	534.5	11.1	859	7	US-11-188-743-16
7	425.5	8.8	769	7	US-11-188-743-15
8	385	8.0	756	7	US-11-188-743-20
9	356	7.4	737	7	US-11-128-420-9
10	355.5	7.4	658	6	US-10-467-657-4782
11	309.5	6.4	389	7	US-11-188-743-23
12	274	5.7	133	7	US-11-128-420-13
13	265.5	5.5	147	7	US-11-128-420-14
14	249	5.2	1151	7	US-11-128-420-10
15	168.5	3.5	1404	6	US-10-878-556A-169
16	165.5	3.4	1960	7	US-11-069-834-50
17	163	3.4	5024	6	US-10-793-626-2964
18	160.5	3.3	619	6	US-10-485-517-374
19	159	3.3	264	7	US-11-188-743-24
20	159	3.3	264	7	US-11-188-743-25
21	153.5	3.2	1960	7	US-11-069-834-48
22	152	3.2	663	7	US-11-196-475-78
23	149	3.1	663	7	US-11-196-475-70
24	148.5	3.1	693	7	US-11-196-475-72
25	147.5	3.1	1976	7	US-11-069-834-52

ALIGNMENTS

RESULT 1

US-11-188-743-17
; Sequence 17, Application US/11188743
; Publication No. US20050272140A1

GENERAL INFORMATION:
; APPLICANT: Nicolaides, Nicholas
; APPLICANT: Sasse, Philip
; APPLICANT: Kinzler, Kenneth
; APPLICANT: Grasso, Luigi
; APPLICANT: Vogelstein, Bert
; TITLE OF INVENTION: Methods for generating hypermutable
; FILE REFERENCE: 01107.00097
; CURRENT APPLICATION NUMBER: US/11/188,743
; CURRENT FILING DATE: 2005-07-26
; PRIOR APPLICATION NUMBER: US/10/641,068
; PRIOR FILING DATE: 2003-08-15
; PRIOR APPLICATION NUMBER: US/09/788,657
; PRIOR FILING DATE: 2001-02-21
; PRIOR APPLICATION NUMBER: 60/184,336
; PRIOR FILING DATE: 2000-02-23
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: FastSeq for Windows Version 3.0

SEQ ID NO 17
; LENGTH: 932
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-188-743-17

Query Match 100.0%; Score 4812; DB 7; Length 932;
Best Local Similarity 100.0%; Pred. No. 6.5e-284;
Matches 932; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MKQLPAATVRLSSSQIITSVVVKELIENSLDAGATSDVVKELNYGFDKIEVRDNGEG	60
Db	1	MKQLPAATVRLSSSQIITSVVVKELIENSLDAGATSDVVKELNYGFDKIEVRDNGEG	60
QY	61	IKADVPAMWKYTSKINSHEDLENLTYYGPRGALSGICIAEVLITRTAADFSTQ	120
Db	61	IKADVPAMWKYTSKINSHEDLENLTYYGPRGALSGICIAEVLITRTAADFSTQ	120
QY	121	YVLDGSHILSQKPSHLGGTGTVALRFLKPLPVKQFYSTAKCKDEIKKTDLLMSFG	180
Db	121	YVLDGSHILSQKPSHLGGTGTVALRFLKPLPVKQFYSTAKCKDEIKKTDLLMSFG	180
QY	181	ILKPLDLRVFVHAKAVIWKQSRVSDHKVALMSVLGTAVMNNMESFYHSESQIYLSGFL	240
Db	181	ILKPLDLRVFVHAKAVIWKQSRVSDHKVALMSVLGTAVMNNMESFYHSESQIYLSGFL	240

```
QY 241 PKCDADHSFTSLSTPERSFIFINSRPHVQKDIKLIRHHYNLKCCLKESTRLYPVPFLKID 300
Db 241 PKCDADHSFTSLSTPERSFIFINSRPHVQKDIKLIRHHYNLKCCLKESTRLYPVPFLKID 300
QY 301 VPTADVNLTPDKSOVLQNKESVLIALENLMTTCYGPLSTNSYENNKTDVSAADIVL 360
Db 301 VPTADVNLTPDKSOVLQNKESVLIALENLMTTCYGPLSTNSYENNKTDVSAADIVL 360
QY 361 SKTAEVDVLFNKVSSGKNYSNVDTSVIPQNDHNDSEKNTDCLNHQISIGDFGYGH 420
Db 361 SKTAEVDVLFNKVSSGKNYSNVDTSVIPQNDHNDSEKNTDCLNHQISIGDFGYGH 420
QY 421 CSSEISNDKNTKNAFODISMSNVSWNSQTEYSKTCFISSVKHTQSENGKNDHDESGE 480
Db 421 CSSEISNDKNTKNAFODISMSNVSWNSQTEYSKTCFISSVKHTQSENGKNDHDESGE 480
QY 481 NEEBAGLENSSEISADEWSRGNILKNSVGENIEPVKILVPEKSLPCKVSNNNYPPEQWN 540
Db 481 NEEBAGLENSSEISADEWSRGNILKNSVGENIEPVKILVPEKSLPCKVSNNNYPPEQWN 540
QY 541 LNEDSCNKKSNVIDNKSQKVTAYDLSNRVIAKKPMSASALFVODHRPQFLIENPKTSLED 600
Db 541 LNEDSCNKKSNVIDNKSQKVTAYDLSNRVIAKKPMSASALFVODHRPQFLIENPKTSLED 600
QY 601 ATLQIEELWKTLSBEEKLYEKKATKDLERYNSQMKRAIEQESQMSLKDGRKKIKPTSAW 660
Db 601 ATLQIEELWKTLSBEEKLYEKKATKDLERYNSQMKRAIEQESQMSLKDGRKKIKPTSAW 660
QY 720 NLAQKHKLKTSLSNQPKDELLOSOIEKRRSQNIKWQIPFSMKNLKINFKKQNKVDLEE 720
Db 720 NLAQKHKLKTSLSNQPKDELLOSOIEKRRSQNIKWQIPFSMKNLKINFKKQNKVDLEE 720
QY 780 KDEPCLIHNLRFPPDAWLTSTKTEVMLNPNRYVEEALLFKRLLENHKLPAEPLEK2IMLTE 780
Db 780 KDEPCLIHNLRFPPDAWLTSTKTEVMLNPNRYVEEALLFKRLLENHKLPAEPLEK2IMLTE 780
QY 840 SLFNGSHVLDVLYKMTADDQRYSGTSLSDPRLTANGFKIKLIPGVSTENYLEIEGMAN 840
Db 840 SLFNGSHVLDVLYKMTADDQRYSGTSLSDPRLTANGFKIKLIPGVSTENYLEIEGMAN 840
QY 900 CLPFFGVADLKEILNAILNRNAKEVYECRPRKVISYLEGEAVRLSRQLPMYLSKEDIQDI 900
Db 900 CLPFFGVADLKEILNAILNRNAKEVYECRPRKVISYLEGEAVRLSRQLPMYLSKEDIQDI 900
QY 932 IYRMKHQFGNEIKECVHGRRPFFHLLTYLPETT 932
Db 932 IYRMKHQFGNEIKECVHGRRPFFHLLTYLPETT 932
```

RESULT 2

```
US-11-188-743-18
; Sequence 18, Application US/11188743
; Publication No. US20050272140A1
; GENERAL INFORMATION:
; APPLICANT: Nicolaides, Nicholas
; APPLICANT: Sasse, Philip
; APPLICANT: Kinzler, Kenneth
; APPLICANT: Grasso, Luigi
; APPLICANT: Vogelstein, Bert
; TITLE OF INVENTION: Methods for generating hypermutable
; TITLE OF INVENTION: Yeast
; FILE REFERENCE: 01107.00097
; CURRENT APPLICATION NUMBER: US/11/188,743
; CURRENT FILING DATE: 2005-07-26
; PRIOR APPLICATION NUMBER: US/10/641,068
; PRIOR FILING DATE: 2003-08-15
; PRIOR APPLICATION NUMBER: US/09/788,657
; PRIOR FILING DATE: 2001-02-21
; PRIOR APPLICATION NUMBER: 60/184,336
; PRIOR FILING DATE: 2000-02-23
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: FastSeq for Windows Version 3.0
```

```
; SEQ ID NO 18
; LENGTH: 932
; TYPE: EXT
; ORGANISM: Homo sapiens
US-11-188-743-18
```

```
Query Match 100.0%; Score 4812; DB 7; Length 932;
Beat Local Similarity 100.0%; Pred. No. 6.5e-284;
Matches 932; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 MKQLPAAATVRLSSQIITSVSVVKELIENSLDAGATSDVKLENYGFDKIEVRDNGEG 60
Db 1 MKQLPAAATVRLSSQIITSVSVVKELIENSLDAGATSDVKLENYGFDKIEVRDNGEG 60
QY 61 IKAVDAPVMAKYYTTSKINSHEDLENLTTYGFRGALGSI CCIABVLITTTAAADFSTQ 120
Db 61 IKAVDAPVMAKYYTTSKINSHEDLENLTTYGFRGALGSI CCIABVLITTTAAADFSTQ 120
QY 121 YVLDGSGHILSQKPSHLGGQTTVTALRLFKNLPVRKQFYSTAKCKDEIKKIQDLLMSFG 180
Db 121 YVLDGSGHILSQKPSHLGGQTTVTALRLFKNLPVRKQFYSTAKCKDEIKKIQDLLMSFG 180
QY 181 ILKPDRLRIVFVHNKAVIWKQSRVSDHKMALMSVLGTAVMNMWESFOYHSEESQIYLSGFL 240
Db 181 ILKPDRLRIVFVHNKAVIWKQSRVSDHKMALMSVLGTAVMNMWESFOYHSEESQIYLSGFL 240
QY 241 PKCDADHSFTSLSTPERSFIFINSRPHVQKDIKLIRHHYNLKCCLKESTRLYPVPFLKID 300
Db 241 PKCDADHSFTSLSTPERSFIFINSRPHVQKDIKLIRHHYNLKCCLKESTRLYPVPFLKID 300
QY 301 VPTADVNLTPDKSOVLQNKESVLIALENLMTTCYGPLSTNSYENNKTDVSAADIVL 360
Db 301 VPTADVNLTPDKSOVLQNKESVLIALENLMTTCYGPLSTNSYENNKTDVSAADIVL 360
QY 361 SKTAEVDVLFNKVSSGKNYSNVDTSVIPQNDHNDSEKNTDCLNHQISIGDFGYGH 420
Db 361 SKTAEVDVLFNKVSSGKNYSNVDTSVIPQNDHNDSEKNTDCLNHQISIGDFGYGH 420
QY 421 CSSEISNDKNTKNAFODISMSNVSWNSQTEYSKTCFISSVKHTQSENGKNDHDESGE 480
Db 421 CSSEISNDKNTKNAFODISMSNVSWNSQTEYSKTCFISSVKHTQSENGKNDHDESGE 480
QY 481 NEEBAGLENSSEISADEWSRGNILKNSVGENIEPVKILVPEKSLPCKVSNNNYPPEQWN 540
Db 481 NEEBAGLENSSEISADEWSRGNILKNSVGENIEPVKILVPEKSLPCKVSNNNYPPEQWN 540
QY 541 LNEDSCNKKSNVIDNKSQKVTAYDLSNRVIAKKPMSASALFVODHRPQFLIENPKTSLED 600
Db 541 LNEDSCNKKSNVIDNKSQKVTAYDLSNRVIAKKPMSASALFVODHRPQFLIENPKTSLED 600
QY 601 ATLQIEELWKTLSBEEKLYEKKATKDLERYNSQMKRAIEQESQMSLKDGRKKIKPTSAW 660
Db 601 ATLQIEELWKTLSBEEKLYEKKATKDLERYNSQMKRAIEQESQMSLKDGRKKIKPTSAW 660
QY 661 NLAQKHKLKTSLSNQPKDELLOSOIEKRRSQNIKWQIPFSMKNLKINFKKQNKVDLEE 720
Db 661 NLAQKHKLKTSLSNQPKDELLOSOIEKRRSQNIKWQIPFSMKNLKINFKKQNKVDLEE 720
QY 721 KDEPCLIHNLRFPPDAWLTSTKTEVMLNPNRYVEEALLFKRLLENHKLPAEPLEK2IMLTE 780
Db 721 KDEPCLIHNLRFPPDAWLTSTKTEVMLNPNRYVEEALLFKRLLENHKLPAEPLEK2IMLTE 780
QY 781 SLFNGSHVLDVLYKMTADDQRYSGTSLSDPRLTANGFKIKLIPGVSTENYLEIEGMAN 840
Db 781 SLFNGSHVLDVLYKMTADDQRYSGTSLSDPRLTANGFKIKLIPGVSTENYLEIEGMAN 840
QY 841 CLPFFGVADLKEILNAILNRNAKEVYECRPRKVISYLEGEAVRLSRQLPMYLSKEDIQDI 900
Db 841 CLPFFGVADLKEILNAILNRNAKEVYECRPRKVISYLEGEAVRLSRQLPMYLSKEDIQDI 900
QY 932 IYRMKHQFGNEIKECVHGRRPFFHLLTYLPETT 932
Db 932 IYRMKHQFGNEIKECVHGRRPFFHLLTYLPETT 932
```

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 21, 2005, 20:20:47 ; Search time 189 Seconds
(without alignments)
2166.672 Million cell updates/sec

Title: US-10-079-429A-4

Perfect score: 932

Sequence: 1 MKQLPATVRLLSSSQIITS.....KECVHGRFFHLLTLPETT 932

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 2443163 seqs, 439378781 residues

Word size: 0

Total number of hits satisfying chosen parameters: 2443163

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : A_Geneseq_21.*

- 1: Geneseq1980s.*
- 2: Geneseq1990s.*
- 3: Geneseq2000s.*
- 4: Geneseq2001s.*
- 5: Geneseq2002s.*
- 6: Geneseq2003as.*
- 7: Geneseq2003bs.*
- 8: Geneseq2004s.*
- 9: Geneseq2005s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	932	100.0	932	4	Aab85852 Human PMS
2	932	100.0	932	4	Aag63953 Amino aci
3	932	100.0	932	4	Aag63954 Amino aci
4	932	100.0	932	5	Aau98776 Human pos
5	932	100.0	932	5	Aae24357 Human mis
6	932	100.0	932	5	Aao18553 Human mis
7	932	100.0	932	5	Aae28277 Human MLH
8	932	100.0	932	5	Aae24684 Human PMS
9	932	100.0	932	6	Abu07972 Human PMS
10	932	100.0	932	6	Abu07971 Human PMS
11	932	100.0	932	6	Abu89659 Human PMS
12	932	100.0	932	6	Abu89660 Human PMS
13	932	100.0	932	6	Abu07415 Human Mut.
14	932	100.0	932	6	Aao27514 Human mis
15	932	100.0	932	6	Aao27515 Human mis
16	932	100.0	932	7	Ada06246 Human mis
17	932	100.0	932	7	Ada06244 Human mis
18	932	100.0	932	7	Adc89605 Human PMS
19	932	100.0	932	7	Adc89607 Human PMS
20	932	100.0	932	7	Adf17892 Human PMS
21	932	100.0	932	7	Adg62881 Human PMS
22	932	100.0	932	7	Adg62882 Human PMS
23	932	100.0	932	7	Adh62629 Human mis
24	932	100.0	932	7	Adh60983 Human mis

ALIGNMENTS

RESULT 1

AAB85852 ID AAB85852 standard; protein; 932 AA.

XX AC AAB85852;

XX DT 29-OCT-2001 (first entry)

XX DE Human PMS1 protein.

XX KW Hypermutable bacteria; mismatch repair gene; MMR gene; MutH; MutS; MutL;

XX KW MutY; PMS2; MLH1; MLH3; PMSR; biocatalysis; bioremediation; biochemical; drug discovery; detoxification; toxin; biotransformation; PMS1.

XX OS Homo sapiens.

XX PN WO200159092-A2.

XX PD 16-AUG-2001.

XX PF 12-FEB-2001; 2001WO-US004339.

XX PR 11-FEB-2000; 2000US-0181929P.

XX PA (UYJO) UNIV JOHNS HOPKINS.

XX PI Nicolaides NC, Sass PM, Grasso L, Vogelstein B, Kinzler KW;

XX DR WPI; 2001-514664/56.

XX DR N-PSDB; AAH76365.

XX PT Making hypermutable bacteria for biocatalysis, bioremediation and drug

XX PT discovery, involves introducing polynucleotide comprising dominant

XX PT negative allele of mismatch repair gene under regulatory sequence

XX PS Example 1; Page 41; 68pp; English.

XX CC The invention provides a method for generating a hypermutable bacteria.

XX CC The method involves introducing a polynucleotide having a dominant

XX CC negative allele of a mismatch repair (MMR) gene under the control of an

XX CC inducible transcription regulatory sequence, into a bacterium. The cell

XX CC becomes inducibly hypermutable. The method is useful to create desirable

XX CC output traits for commercial applications, using dominant negative

XX CC alleles of mismatch repair proteins. The mismatch repair gene is a MutH,

XX CC MutS, MutL or MutY homologue and can be selected from PMS2, MLH1, MLH3,

XX CC PMSR or PMSR homologue. The hypermutable bacteria is useful for the

XX CC

CC production, biocatalysis, bioremediation and drug discovery. It is also
CC useful in manufacturing industry for the generation of new biochemicals
CC useful for detoxifying noxious chemicals from by-products of
CC manufacturing processes or those used as catalysts, for remediation of
CC toxins present in the environment including polychlorobenzenes, heavy
CC metals and other environmental hazards for which there is a need to
CC remove them from the environment. The hypermutable bacteria is also
CC useful for screening novel mutations in a gene or a set of genes that
CC produce variant siblings that exhibit a new output trait not found in
CC wild type cells. The bacteria are also useful for producing increased
CC quantity or quality of protein or non-protein therapeutic molecule e.g.
CC Penicillin G, Erythromycin and Clavulanic acid, by biotransformation.
CC Dominant negative alleles of the MMR gene are useful for producing higher
CC quantities of recombinant polypeptides. The present sequence represents a
CC human PMS1 protein
XX
SQ Sequence 932 AA;

Query Match 100.0%; Score 932; DB 4; Length 932;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 932; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKQLPAATVRLSSQIIITSVSVVKELIENSLDAGATSDVKLENYGDKIEVRDNGSG 60
DB 1 MKQLPAATVRLSSQIIITSVSVVKELIENSLDAGATSDVKLENYGDKIEVRDNGSG 60

QY 61 IKAVDAPVMAMKYTTSKINSHEDLENLTYYGFRGEALGSI CCAEVLITTRTAADNFSTQ 120
DB 61 IKAVDAPVMAMKYTTSKINSHEDLENLTYYGFRGEALGSI CCAEVLITTRTAADNFSTQ 120

QY 121 YVLDGSGHLSQKPSHLGGTIVTALRFLKPLVPRKQFYSTAKCKDEIKKIQDLMSFG 180
DB 121 YVLDGSGHLSQKPSHLGGTIVTALRFLKPLVPRKQFYSTAKCKDEIKKIQDLMSFG 180

QY 181 ILKPLDLRIVFVNKAVIWKSRVSDHKMALMSVLGTAVNNMESFOYHSEESQIYLSGFL 240
DB 181 ILKPLDLRIVFVNKAVIWKSRVSDHKMALMSVLGTAVNNMESFOYHSEESQIYLSGFL 240

QY 241 PKCDADHSTSLSTPERSIFINSRPVHQKDILKIRHYNLKLKESTRLPVFPLKID 300
DB 241 PKCDADHSTSLSTPERSIFINSRPVHQKDILKIRHYNLKLKESTRLPVFPLKID 300

QY 301 VPTADVNLITPDKSOVLQNKESVLIALENLMTTCYGPLPSTNSYNNKTDVSAADIVL 360
DB 301 VPTADVNLITPDKSOVLQNKESVLIALENLMTTCYGPLPSTNSYNNKTDVSAADIVL 360

QY 361 SKTAETDVLFNKVESGKNYSNVDTSVIPQNDMHNDESGKNTDDCLNHQISIGDFGYGH 420
DB 361 SKTAETDVLFNKVESGKNYSNVDTSVIPQNDMHNDESGKNTDDCLNHQISIGDFGYGH 420

QY 421 CSSEISNIDKNTKNAFQDISMSNVSWNSQTEYSKTCFISSVKHQTSQENKDKHIDESGE 480
DB 421 CSSEISNIDKNTKNAFQDISMSNVSWNSQTEYSKTCFISSVKHQTSQENKDKHIDESGE 480

QY 481 NEEEAGLENSSEISADEWSEGNILKNSVGENIEPVKILVPEKSLPCKVSNNNYPIPEQWN 540
DB 481 NEEEAGLENSSEISADEWSEGNILKNSVGENIEPVKILVPEKSLPCKVSNNNYPIPEQWN 540

QY 541 LNEДСNKNSVINDKSGKVTAYDLSNVRVKKPMSASALFVQDHRPQFLIENPKTSLED 600
DB 541 LNEДСNKNSVINDKSGKVTAYDLSNVRVKKPMSASALFVQDHRPQFLIENPKTSLED 600

QY 601 ATLQIEELWKTLSSEKLYEEXATKDLERYNSQMKRAIEQESQMSLKGRKKIKTSAW 660
DB 601 ATLQIEELWKTLSSEKLYEEXATKDLERYNSQMKRAIEQESQMSLKGRKKIKTSAW 660

QY 661 NLAQKHKLTSLSNQPKLDELLOSQIEKRSQNIKWQIPFSSMKNLKINFKKQNKVDLEE 720
DB 661 NLAQKHKLTSLSNQPKLDELLOSQIEKRSQNIKWQIPFSSMKNLKINFKKQNKVDLEE 720

QY 721 KQBPCLIHNLRRPDAWLTSTKTEVMLLNPYRVEALLFKRLLENHKLPAEPLKPTMLTE 780
DB 721 KQBPCLIHNLRRPDAWLTSTKTEVMLLNPYRVEALLFKRLLENHKLPAEPLKPTMLTE 780

QY 781 SLFNGSHYLDVLYKMTADDQRYSGSTYLSDDRLTANGFKIKLIPGVSTITENYLEIGMAN 840
DB 781 SLFNGSHYLDVLYKMTADDQRYSGSTYLSDDRLTANGFKIKLIPGVSTITENYLEIGMAN 840

QY 841 CLPFGYGVADLKEILNAILNRNAKEVYECRPRKVISYLEGEAVRLSRQLPMLSKEIDIQI 900
DB 841 CLPFGYGVADLKEILNAILNRNAKEVYECRPRKVISYLEGEAVRLSRQLPMLSKEIDIQI 900

QY 901 IYRMKHQFGNEIKECVHGRRPFPHLTYLPEIT 932
DB 901 IYRMKHQFGNEIKECVHGRRPFPHLTYLPEIT 932

RESULT 2
AAG63953
ID AAG63953 standard; protein; 932 AA.
XX
AC AAG63953;
XX
DT 29-OCT-2001 (first entry)
XX
DE Amino acid sequence of human mismatch repair protein PMS2.
XX
KW PMS2; mismatch repair gene; MMR gene; hypermutable yeast.
XX
OS Homo sapiens.
XX
FN WO200162945-A1.
XX
PD 30-AUG-2001.
XX
PF 21-FEB-2001; 2001WO-US005447.
XX
PR 23-FEB-2000; 2000US-0184336P.
XX
PA (UYJO) UNIV JOHNS HOPKINS.
PA (NICO/) NICOLAIDES N C.
PA (SASS/) SASS P M.
PA (GRAS/) GRASSO L.
PA (VOGE/) VOGELSTEIN B.
PA (KINZ/) KINZLER K W.
XX
PI Nicolaides NC, Sass PM, Grasso L, Vogelstein B, Kinzler KW;
XX
XX WPI; 2001-522820/57.
DR N-PSDB; AAH75041.
XX
XX Making hypermutable yeast that exhibit novel selected output traits for
PT commercial applications, comprises introducing polynucleotide containing
PT dominant negative allele of mismatch repair gene.
XX
XX Disclosure; Page 38; 60pp; English.
PS
CC The present sequence represents human PMS2. PMS2 is a mismatch repair
CC (MMR) gene. The specification describes a method for making a
CC hypermutable yeast, comprising introducing a polynucleotide containing a
CC dominant negative allele of a mismatch repair (MMR) gene, into a yeast,
CC whereby the cell becomes hypermutable. The method is useful to create
CC desirable output traits for commercial applications, using dominant
CC negative alleles of mismatch repair proteins. The hypermutable yeast is
CC useful for production, biocatalysis, bioremediation and drug discovery.
CC It is also useful in genetic screens for the direct selection of variant
CC subclones that exhibit new output traits. The hypermutable yeast is also
CC useful in the manufacturing industry for the generation of new
CC biochemicals, for detoxifying noxious chemicals from by-products of
CC manufacturing processes or those used as catalysts, for remediation of
CC toxins present in the environment including polychlorobenzenes, heavy
CC metals and other environmental hazards for which there is a need to
CC remove them from the environment. The yeast is also useful for producing
CC increased quantity or quality of protein or non-protein therapeutic
CC molecule e.g., Penicillin G, Erythromycin and Clavulanic acid, by
CC biotransformation

Gencore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 21, 2005, 20:24:48 ; Search time 48 Seconds

(without alignments)
1605.286 Million cell updates/sec

Title: US-10-079-429A-4

Perfect score: 932

Sequence: 1 MKQLPAATVRLSSQIITS.....KSCVHGPPFFHLTYLPEPT 932

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 572060 seqs, 82675679 residues

Word size : 0

Total number of hits satisfying chosen parameters: 572060

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued Patents AA:*

- 1: /cgn2_6/prodata/1/iaa/5 COMB.pep.*
- 2: /cgn2_6/prodata/1/iaa/6 COMB.pep.*
- 3: /cgn2_6/prodata/1/iaa/H COMB.pep.*
- 4: /cgn2_6/prodata/1/iaa/PCUS COMB.pep.*
- 5: /cgn2_6/prodata/1/iaa/RE COMB.pep.*
- 6: /cgn2_6/prodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	932	100.0	932	2	US-08-294-312B-4
2	932	100.0	932	2	US-08-468-024B-4
3	932	100.0	932	2	US-09-708-200-11
4	932	100.0	932	2	US-08-465-679-4
5	932	100.0	932	2	US-09-788-657-17
6	932	100.0	932	2	US-09-788-657-18
7	932	100.0	932	2	US-09-712-691-9
8	932	100.0	932	2	US-10-641-068-17
9	932	100.0	932	2	US-10-641-068-18
10	912	97.9	921	2	US-09-349-016-7787
11	893	95.8	932	2	US-09-707-468C-9
12	133	14.3	133	2	US-09-788-657-21
13	133	14.3	133	2	US-09-712-691-15
14	133	14.3	133	2	US-10-641-068-21
15	16	1.7	580	2	US-09-198-452A-870
16	16	1.7	594	2	US-09-438-185A-814
17	12	1.3	147	2	US-09-749-601A-14
18	12	1.3	426	2	US-08-676-444-40
19	12	1.3	779	2	US-09-749-601A-12
20	9	1.0	64	1	US-08-209-521-19
21	9	1.0	64	1	US-08-209-521-20
22	9	1.0	64	2	US-08-861-810-129
23	9	1.0	64	2	US-08-861-810-130
24	9	1.0	64	2	US-08-352-902D-129
25	9	1.0	64	2	US-08-352-902D-130
26	9	1.0	64	2	US-09-265-503B-129
27	9	1.0	64	2	US-09-265-503B-130

28	9	1.0	133	2	US-09-708-200-17	Sequence 17, Appl
29	9	1.0	133	2	US-09-707-468C-15	Sequence 15, Appl
30	9	1.0	133	2	US-09-749-601A-13	Sequence 13, Appl
31	9	1.0	158	2	US-09-489-039A-8515	Sequence 8515, Ap
32	9	1.0	244	2	US-09-248-796A-19147	Sequence 19147, A
33	9	1.0	264	2	US-09-788-657-24	Sequence 24, Appl
34	9	1.0	264	2	US-09-788-657-25	Sequence 25, Appl
35	9	1.0	264	2	US-10-641-068-24	Sequence 24, Appl
36	9	1.0	264	2	US-10-641-068-25	Sequence 25, Appl
37	9	1.0	361	1	US-08-209-521-5	Sequence 5, Appl
38	9	1.0	361	2	US-08-961-810-1	Sequence 1, Appl
39	9	1.0	361	2	US-08-352-902D-1	Sequence 1, Appl
40	9	1.0	361	2	US-09-265-503B-1	Sequence 1, Appl
41	9	1.0	389	2	US-09-788-657-23	Sequence 23, Appl
42	9	1.0	389	2	US-10-641-068-23	Sequence 23, Appl
43	9	1.0	511	2	US-08-676-444-42	Sequence 42, Appl
44	9	1.0	615	2	US-08-676-444-44	Sequence 44, Appl
45	9	1.0	674	2	US-09-543-681A-5715	Sequence 5715, Ap

ALIGNMENTS

RESULT 1

US-08-294-312B-4
; Sequence 4, Application US/08294312B
; Patent No. 6380369
; GENERAL INFORMATION:
; APPLICANT: Adams et al.
; TITLE OF INVENTION: Human DNA Mismatch Repair Proteins
; FILE REFERENCE: PFI06P2
; CURRENT APPLICATION NUMBER: US/08/294,312B
; CURRENT FILING DATE: 1994-08-23
; PRIOR APPLICATION NUMBER: 08/210,143
; PRIOR FILING DATE: 1994-03-16
; PRIOR APPLICATION NUMBER: 08/187,757
; PRIOR FILING DATE: 1994-01-27
; NUMBER OF SEQ ID NOS: 78
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 4
; LENGTH: 932
; TYPE: PRT
; ORGANISM: homo sapiens
; US-08-294-312B-4

Query Match	100.0%;	Score	932;	DB	2;	Length	932;
Best Local Similarity	100.0%;	Pred. No.	0;				
Matches	932;	Conservative	0;	Mismatches	0;	Indels	0;
Gaps	0;						
QY	1	MKQLPAATVRLSSQIITSVVVKELIENSIDAGATSV	VDVKLENYGFDKIEVRDNGEG	60			
Db	1	MKQLPAATVRLSSQIITSVVVKELIENSIDAGATSV	VDVKLENYGFDKIEVRDNGEG	60			
QY	61	IXADAPVAMKYTTSKINSHEDLENLT	TYGFRGEALGSGICCAEVLITR	TAADNFSTQ	120		
Db	61	IXADAPVAMKYTTSKINSHEDLENLT	TYGFRGEALGSGICCAEVLITR	TAADNFSTQ	120		
QY	121	YVLDGSGHTLSQKPSHLGGTTVALR	FKNLPVRKQFYSTAKCKDEIKKI	ODLLMSFG	180		
Db	121	YVLDGSGHTLSQKPSHLGGTTVALR	FKNLPVRKQFYSTAKCKDEIKKI	ODLLMSFG	180		
QY	181	ILKPLDIRIVFNKAVIWKQSRVSDHKMALMSV	LGTAVMNNMESFQVHSESIYLSGFL	240			
Db	181	ILKPLDIRIVFNKAVIWKQSRVSDHKMALMSV	LGTAVMNNMESFQVHSESIYLSGFL	240			
QY	241	PKCDADHSFTSLSTPERSPIFINSR	VPVHOKDILKLRHHYNLCKLKESTR	LYPVFFLKID	300		
Db	241	PKCDADHSFTSLSTPERSPIFINSR	VPVHOKDILKLRHHYNLCKLKESTR	LYPVFFLKID	300		
QY	301	VPTADVNLTPDKSOVLQNKESVLIALENL	MTTCVGPLPSTNSYENNKTDVSAADIVL	360			
Db	301	VPTADVNLTPDKSOVLQNKESVLIALENL	MTTCVGPLPSTNSYENNKTDVSAADIVL	360			

361 SKTAETDVLFNKVESGKNSVNDTSVIPPONDMDHNDSEKNTDCLNHQISIGDFGYGH 420
361 SKTAETDVLFNKVESGKNSVNDTSVIPPONDMDHNDSEKNTDCLNHQISIGDFGYGH 420
421 CSSEISNIDKNTKNAFQDISMSNVSWNSQTEYSKTCFISVVKHTQSENGKNDHIDESGE 480
421 CSSEISNIDKNTKNAFQDISMSNVSWNSQTEYSKTCFISVVKHTQSENGKNDHIDESGE 480
481 NEEAAGLENSSEISADEWSRGNIKNSVGENIPEVKILVPEKSLPKCKVSNNNYPIPEQWN 540
481 NEEAAGLENSSEISADEWSRGNIKNSVGENIPEVKILVPEKSLPKCKVSNNNYPIPEQWN 540
541 LNEEDSCNKSNSVNDKSGKVTAYDLLSNRVIKKPMASALFVQDHRPQFLIENPKTSLED 600
541 LNEEDSCNKSNSVNDKSGKVTAYDLLSNRVIKKPMASALFVQDHRPQFLIENPKTSLED 600
601 ATLQIEELWKTLSSEBKLYEEKATKDLERYNSQMKRAIEQESQMSLKDGGRKKIKPTSAS 660
601 ATLQIEELWKTLSSEBKLYEEKATKDLERYNSQMKRAIEQESQMSLKDGGRKKIKPTSAS 660
661 NLAQKHKLKTSLSNQPKLDELLOSOIEKRRSONIKMVQIPFMSMKNLKNFKKQKNDVLEE 720
661 NLAQKHKLKTSLSNQPKLDELLOSOIEKRRSONIKMVQIPFMSMKNLKNFKKQKNDVLEE 720
721 KQEPCLIHNLRFPPDAMLTSKTEVMLNPNRYVEEALLFKRLLENHKLPAEPLKPIMLTE 780
721 KQEPCLIHNLRFPPDAMLTSKTEVMLNPNRYVEEALLFKRLLENHKLPAEPLKPIMLTE 780
781 SLFNGSHYLDVLYKMTADDQRYSGSTYLSDPRLTANGFKIKLIPGVSIITENYLEIEGMA 840
781 SLFNGSHYLDVLYKMTADDQRYSGSTYLSDPRLTANGFKIKLIPGVSIITENYLEIEGMA 840
841 CLPFGYGVADLKEILNALNNAKEVEYECRPRKVISYLEGEAVRLSRQLPMYLSKEDIQDI 900
841 CLPFGYGVADLKEILNALNNAKEVEYECRPRKVISYLEGEAVRLSRQLPMYLSKEDIQDI 900
901 IYRMKHQFGNEIKECVHGRPFPHHLYLPETT 932
901 IYRMKHQFGNEIKECVHGRPFPHHLYLPETT 932

RESULT 2

US-08-468-024B-4
; Sequence 4, Application US/08468024B
; Patent No. 6416984
; GENERAL INFORMATION:
; APPLICANT: Haseltine et al.
; TITLE OF INVENTION: Human DNA Mismatch Repair Proteins
; FILE REFERENCE: PP106P3
; CURRENT APPLICATION NUMBER: US/08/468,024B
; PRIOR FILING DATE: 1995-06-06
; PRIOR APPLICATION NUMBER: 08/294,312
; PRIOR FILING DATE: 1994-08-23
; PRIOR APPLICATION NUMBER: 08/210,143
; PRIOR FILING DATE: 1994-03-16
; PRIOR APPLICATION NUMBER: 08/187,757
; PRIOR FILING DATE: 1994-01-27
; NUMBER OF SEQ ID NOS: 78
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4
; LENGTH: 932
; TYPE: PRT
; ORGANISM: homo sapiens
US-08-468-024B-4

Query Match 100.0%; Score 932; DB 2; Length 932;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 932; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 MKQLPAATVRLSSQIITSVSVVKELIENSLDAGATSDVVKLENYGFDKIEVRDNGEG 60
1 MKQLPAATVRLSSQIITSVSVVKELIENSLDAGATSDVVKLENYGFDKIEVRDNGEG 60

61 IKAVDAPVMAMKYTTSKINSHEDLENLTGYPRGEALGSIICIAEVLITRTRAADNFSIQ 120
61 IKAVDAPVMAMKYTTSKINSHEDLENLTGYPRGEALGSIICIAEVLITRTRAADNFSIQ 120
121 YVLDGSGHILSQKPSHLGGTIVTALRFPKNIPLVRKQFYSTAKCKDEIKKIQDILLMSGF 180
121 YVLDGSGHILSQKPSHLGGTIVTALRFPKNIPLVRKQFYSTAKCKDEIKKIQDILLMSGF 180
181 ILKPDRLRIVFVANKAVIWKSRVSDHKMALMSVLGTAVNNNMESFOYHSEESQIYLSGFL 240
181 ILKPDRLRIVFVANKAVIWKSRVSDHKMALMSVLGTAVNNNMESFOYHSEESQIYLSGFL 240
241 PKCDADHSFTSLSTPERSFIFINSRPVHQKDIKLIIRHYNLKLCKESTRLYPVFFLKID 300
241 PKCDADHSFTSLSTPERSFIFINSRPVHQKDIKLIIRHYNLKLCKESTRLYPVFFLKID 300
301 VPTADVNLTPDKSQVLLQNKESVLLIENLMTTCYGLPSTNSYNNKTDVSAADIVL 360
301 VPTADVNLTPDKSQVLLQNKESVLLIENLMTTCYGLPSTNSYNNKTDVSAADIVL 360
361 SKTAETDVLFNKVESGKNSVNDTSVIPPONDMDHNDSEKNTDCLNHQISIGDFGYGH 420
361 SKTAETDVLFNKVESGKNSVNDTSVIPPONDMDHNDSEKNTDCLNHQISIGDFGYGH 420
421 CSSEISNIDKNTKNAFQDISMSNVSWNSQTEYSKTCFISVVKHTQSENGKNDHIDESGE 480
421 CSSEISNIDKNTKNAFQDISMSNVSWNSQTEYSKTCFISVVKHTQSENGKNDHIDESGE 480
481 NEEAAGLENSSEISADEWSRGNIKNSVGENIPEVKILVPEKSLPKCKVSNNNYPIPEQWN 540
481 NEEAAGLENSSEISADEWSRGNIKNSVGENIPEVKILVPEKSLPKCKVSNNNYPIPEQWN 540
541 LNEEDSCNKSNSVNDKSGKVTAYDLLSNRVIKKPMASALFVQDHRPQFLIENPKTSLED 600
541 LNEEDSCNKSNSVNDKSGKVTAYDLLSNRVIKKPMASALFVQDHRPQFLIENPKTSLED 600
601 ATLQIEELWKTLSSEBKLYEEKATKDLERYNSQMKRAIEQESQMSLKDGGRKKIKPTSAS 660
601 ATLQIEELWKTLSSEBKLYEEKATKDLERYNSQMKRAIEQESQMSLKDGGRKKIKPTSAS 660
661 NLAQKHKLKTSLSNQPKLDELLOSOIEKRRSONIKMVQIPFMSMKNLKNFKKQKNDVLEE 720
661 NLAQKHKLKTSLSNQPKLDELLOSOIEKRRSONIKMVQIPFMSMKNLKNFKKQKNDVLEE 720
721 KQEPCLIHNLRFPPDAMLTSKTEVMLNPNRYVEEALLFKRLLENHKLPAEPLKPIMLTE 780
721 KQEPCLIHNLRFPPDAMLTSKTEVMLNPNRYVEEALLFKRLLENHKLPAEPLKPIMLTE 780
781 SLFNGSHYLDVLYKMTADDQRYSGSTYLSDPRLTANGFKIKLIPGVSIITENYLEIEGMA 840
781 SLFNGSHYLDVLYKMTADDQRYSGSTYLSDPRLTANGFKIKLIPGVSIITENYLEIEGMA 840
841 CLPFGYGVADLKEILNALNNAKEVEYECRPRKVISYLEGEAVRLSRQLPMYLSKEDIQDI 900
841 CLPFGYGVADLKEILNALNNAKEVEYECRPRKVISYLEGEAVRLSRQLPMYLSKEDIQDI 900
901 IYRMKHQFGNEIKECVHGRPFPHHLYLPETT 932
901 IYRMKHQFGNEIKECVHGRPFPHHLYLPETT 932

RESULT 3

US-09-708-200-11
; Sequence 11, Application US/09708200
; Patent No. 6576468
; GENERAL INFORMATION:
; APPLICANT: Nicolaides, Nicholas C
; APPLICANT: Grasso, Luigi
; APPLICANT: Sasse, Philip M
; TITLE OF INVENTION: METHODS FOR ISOLATING NOVEL ANTIMICROBIAL AGENTS FROM
; TITLE OF INVENTION: HYPERMUTABLE CELLS
; FILE REFERENCE: MOR-0005
; CURRENT APPLICATION NUMBER: US/09/708,200

GenCore version 5.1.6
Copyright (C) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 21, 2005, 20:32:28 ; Search time 168 Seconds
(without alignments)
2317.958 Million cell updates/sec

Title: US-10-079-429A-4
Perfect score: 932
Sequence: 1 MKQLPAATVRLSSSQIITS.....KECVHGRPFPHLTLYLPEPT 932

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 1867569 seqs, 417829326 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1867569

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Published Applications AA Main:
1: /cgn2_6/prodata/1/pubpaa/US07_PUBCOMB.pap:*
2: /cgn2_6/prodata/1/pubpaa/US08_PUBCOMB.pap:*
3: /cgn2_6/prodata/1/pubpaa/US09_PUBCOMB.pap:*
4: /cgn2_6/prodata/1/pubpaa/US10A_PUBCOMB.pap:*
5: /cgn2_6/prodata/1/pubpaa/US10B_PUBCOMB.pap:*
6: /cgn2_6/prodata/1/pubpaa/US11_PUBCOMB.pap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	932	100.0	932	3	US-09-788-657-17
2	932	100.0	932	3	US-09-788-657-18
3	932	100.0	932	3	US-09-912-697-6
4	932	100.0	932	3	US-09-912-697-8
5	932	100.0	932	3	US-09-760-285-18
6	932	100.0	932	4	US-10-079-429-4
7	932	100.0	932	4	US-10-270-839-27
8	932	100.0	932	4	US-10-270-839-29
9	932	100.0	932	4	US-10-243-130-9
10	932	100.0	932	4	US-10-371-857-3
11	932	100.0	932	4	US-10-371-857-4
12	932	100.0	932	4	US-10-371-634-5
13	932	100.0	932	4	US-10-371-634-7
14	932	100.0	932	4	US-10-348-074-3
15	932	100.0	932	4	US-10-348-074-5
16	932	100.0	932	4	US-10-369-845-11
17	932	100.0	932	4	US-10-641-068-17
18	932	100.0	932	4	US-10-641-068-18
19	932	100.0	932	4	US-10-408-765A-481
20	932	100.0	932	4	US-10-813-502-9
21	932	100.0	932	4	US-10-714-228-2
22	932	100.0	932	4	US-10-714-228-4
23	932	100.0	932	5	US-10-933-034-2
24	932	100.0	932	5	US-10-901-650-9
25	932	100.0	932	6	US-11-056-776-13
26	893	95.8	932	5	US-10-850-370-9
27	133	14.3	133	3	US-09-788-657-21

28	133	14.3	133	3	US-09-912-697-14	Sequence 14, Appl
29	133	14.3	133	3	US-09-760-285-24	Sequence 24, Appl
30	133	14.3	133	4	US-10-270-839-35	Sequence 35, Appl
31	133	14.3	133	4	US-10-371-857-12	Sequence 12, Appl
32	133	14.3	133	4	US-10-371-634-13	Sequence 13, Appl
33	133	14.3	133	4	US-10-348-074-11	Sequence 11, Appl
34	133	14.3	133	4	US-10-641-068-21	Sequence 21, Appl
35	133	14.3	133	4	US-10-714-228-6	Sequence 6, Appl
36	16	1.7	580	4	US-10-289-762-870	Sequence 870, App
37	16	1.7	580	4	US-10-282-122A-54985	Sequence 54985, A
38	13	1.4	724	3	US-09-954-950-2	Sequence 2, Appl
39	13	1.4	724	3	US-10-270-839-59	Sequence 59, Appl
40	13	1.4	724	4	US-10-714-228-50	Sequence 50, Appl
41	13	1.4	1141	4	US-10-437-963-127061	Sequence 127061,
42	12	1.3	133	4	US-10-714-228-38	Sequence 38, Appl
43	12	1.3	133	5	US-10-933-034-38	Sequence 38, Appl
44	12	1.3	147	3	US-09-749-601A-14	Sequence 14, Appl
45	12	1.3	779	3	US-09-749-601A-12	Sequence 12, Appl

ALIGNMENTS

RESULT 1
US-09-788-657-17
; Sequence 17, Application US/09788657
; Patent No. US20020123149A1
; GENERAL INFORMATION:
; APPLICANT: Nicolaides, Nicholas
; APPLICANT: Sasse, Philip
; APPLICANT: Kinzler, Kenneth
; APPLICANT: Grasso, Luigi
; APPLICANT: Vogelstein, Bert
; TITLE OF INVENTION: Methods for generating hypermutable
; FILE OF INVENTION: Yeast
; FILE REFERENCE: 01107.00097
; CURRENT APPLICATION NUMBER: US/09/788,657
; CURRENT FILING DATE: 2001-02-21
; PRIOR APPLICATION NUMBER: 60/184,336
; PRIOR FILING DATE: 2000-02-23
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 17
; LENGTH: 932
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-788-657-17

Query Match		100.0%;	Score 932;	DB 3;	Length 932;
Best Local Similarity		100.0%;	Pred. No. 0;		
Matches 932;		Conservative	0;	Mismatches	0;
				Indels	0;
				Gaps	0;
QY	1	MKQLPAATVRLSSSQIITSVVVKELIENSLDAGATSDVVKLENYGFDPKIEVRN	GEG	60	
DB	1	MKQLPAATVRLSSSQIITSVVVKELIENSLDAGATSDVVKLENYGFDPKIEVRN	GEG	60	
QY	61	IKADAPVAMKYTTSKINSHEDLENLTYYGFRGALSGICCAIEVLITRTAADNFSTQ	120		
DB	61	IKADAPVAMKYTTSKINSHEDLENLTYYGFRGALSGICCAIEVLITRTAADNFSTQ	120		
QY	121	VYLDGSGHILSKPQSHLGQGTVTALRFLPNLVRKQFYSTAKKCKDEIKKIIDLMSFG	180		
DB	121	VYLDGSGHILSKPQSHLGQGTVTALRFLPNLVRKQFYSTAKKCKDEIKKIIDLMSFG	180		
QY	181	ILKPDRLRVFVNKAVIWKSRVSDHKMALMSVLGTAVNMNMFQYHSEESQIYLSGFL	240		
DB	181	ILKPDRLRVFVNKAVIWKSRVSDHKMALMSVLGTAVNMNMFQYHSEESQIYLSGFL	240		
QY	241	PXCDADHSFTSLSTPERSFIFINSRPVHOKDILKIRHYNLKLKESTRLYPVFFLKID	300		
DB	241	PXCDADHSFTSLSTPERSFIFINSRPVHOKDILKIRHYNLKLKESTRLYPVFFLKID	300		
QY	301	VPTADVNLTPDKSQVLLQNKESVLIALLENLMTTCYGPLPSTNSYENNKTVSAADIVL	360		

Db	301	VPTADVNLTPDKSOVLQNKESVLI	IALENIMTTTCYGLPSTNSYNNKTDVSAADIVL	360
Qy	361	SKTAETDVLFNKVBSSGKNYSNVDTSV	IPFQNDMENDESCKNTDCLNHQISIGDFGYGH	420
Db	361	SKTAETDVLFNKVBSSGKNYSNVDTSV	IPFQNDMENDESCKNTDCLNHQISIGDFGYGH	420
Qy	421	CSSEISNIDKNTKNAFQDISMSNVSWNS	QTSYKTCFISSVKHTQSENGKNDHIDESGE	480
Db	421	CSSEISNIDKNTKNAFQDISMSNVSWNS	QTSYKTCFISSVKHTQSENGKNDHIDESGE	480
Qy	481	NEEEAGLENSSEISADEWSRGNILKNS	VGENIEPVKILVPEKSLPCKVSNNNYP	540
Db	481	NEEEAGLENSSEISADEWSRGNILKNS	VGENIEPVKILVPEKSLPCKVSNNNYP	540
Qy	541	LNEDSCNKKSNVIDNKS	GVKTAYDILLSNRVIKKPMSASALFVQDHRPQFLIENPKTSLED	600
Db	541	LNEDSCNKKSNVIDNKS	GVKTAYDILLSNRVIKKPMSASALFVQDHRPQFLIENPKTSLED	600
Qy	601	ATLOIEELWKTLSSEBEKLYE	EKATKDLERYNSQMKRAIEQBSQMSLKDGGRKKIKPTSAW	660
Db	601	ATLOIEELWKTLSSEBEKLYE	EKATKDLERYNSQMKRAIEQBSQMSLKDGGRKKIKPTSAW	660
Qy	661	NLAQKHKLKTSLSNQPKLDELLOSOI	EKRBSQNIKWQIPFSMKNLKINFKKQNKVDLEE	720
Db	661	NLAQKHKLKTSLSNQPKLDELLOSOI	EKRBSQNIKWQIPFSMKNLKINFKKQNKVDLEE	720
Qy	721	KDEPCLIHNLRPDPDAWMTSKTEV	MLNPNRYVEEALLFKRLLENHKLPAEPLKPIMLTE	780
Db	721	KDEPCLIHNLRPDPDAWMTSKTEV	MLNPNRYVEEALLFKRLLENHKLPAEPLKPIMLTE	780
Qy	781	SLFNGSHYLDVLYKMTADDQRYSG	STYLSDPRLTANGFKIKLIPGVSITENYLEIEGMA	840
Db	781	SLFNGSHYLDVLYKMTADDQRYSG	STYLSDPRLTANGFKIKLIPGVSITENYLEIEGMA	840
Qy	841	CLPFFYGVADLKELNALNLRNAKEV	YECRPRKVISYLEGEAVRLSRQLPMLSKEIDIQDI	900
Db	841	CLPFFYGVADLKELNALNLRNAKEV	YECRPRKVISYLEGEAVRLSRQLPMLSKEIDIQDI	900
Qy	901	IYRMKHQFGNEIKECVHGRPFPHLT	YLPETT 932	
Db	901	IYRMKHQFGNEIKECVHGRPFPHLT	YLPETT 932	
RESULT 3				
US-09-788-657-18				
; Sequence 18, Application US/09788657				
; Patent No. US20020123149A1				
; GENERAL INFORMATION:				
; APPLICANT: Nicolaides, Nicholas				
; APPLICANT: Sassi, Philip				
; APPLICANT: Kinzler, Kenneth				
; APPLICANT: Grasso, Luigi				
; APPLICANT: Vogelstein, Bert				
; TITLE OF INVENTION: Methods for generating hypermutable				
; FILE OF INVENTION: yeast				
; FILE REFERENCE: 01107.00097				
; CURRENT APPLICATION NUMBER: US/09/788,657				
; CURRENT FILING DATE: 2001-02-21				
; PRIOR APPLICATION NUMBER: 60/184,336				
; PRIOR FILING DATE: 2000-02-23				
; NUMBER OF SEQ ID NOS: 25				
; SOFTWARE: FastSeq for Windows Version 3.0				
; SEQ ID NO 18				
; LENGTH: 932				
; TYPE: PRT				
; ORGANISM: Homo sapiens				
US-09-788-657-18				
Query Match 100.0%; Score 932; DB 3; Length 932;				
Best Local Similarity 100.0%; Pred No. 0;				
Matches 932; Conservative 0; Mismatches 0; Indels 0; Gaps 0;				

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 21, 2005, 20:33:19 ; Search time 13 Seconds

(without alignments)
511.317 Million cell updates/sec

Title: US-10-079-429A-4
Perfect score: 932
Sequence: 1 MKQLPAATVRLSSQIITS.....KECVHGRPFHHLTYLPETT 932

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 53982 seqs, 7132107 residues

Word size : 0

Total number of hits satisfying chosen parameters: 53982

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : Published Applications AA New:
1: /cgn2_6/prodata/2/pubpaa/US08_NEW_PUB.pep:
2: /cgn2_6/prodata/2/pubpaa/US06_NEW_PUB.pep:
3: /cgn2_6/prodata/2/pubpaa/US07_NEW_PUB.pep:
4: /cgn2_6/prodata/2/pubpaa/PCT_NEW_PUB.pep:
5: /cgn2_6/prodata/2/pubpaa/US09_NEW_PUB.pep:
6: /cgn2_6/prodata/2/pubpaa/US10_NEW_PUB.pep:
7: /cgn2_6/prodata/2/pubpaa/US11_NEW_PUB.pep:
8: /cgn2_6/prodata/2/pubpaa/US60_NEW_PUB.pep:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	932	100.0	932	US-11-188-743-17	Sequence 17, Appl
2	932	100.0	932	US-11-188-743-18	Sequence 18, Appl
3	133	14.3	133	US-11-188-743-21	Sequence 21, Appl
4	12	1.3	147	US-11-128-420-14	Sequence 14, Appl
5	12	1.3	779	US-11-128-420-12	Sequence 12, Appl
6	9	1.0	133	US-11-128-420-13	Sequence 13, Appl
7	9	1.0	264	US-11-188-743-24	Sequence 24, Appl
8	9	1.0	264	US-11-188-743-25	Sequence 25, Appl
9	9	1.0	389	US-11-188-743-23	Sequence 23, Appl
10	9	1.0	756	US-11-188-743-20	Sequence 20, Appl
11	9	1.0	769	US-11-188-743-15	Sequence 15, Appl
12	9	1.0	862	US-11-128-420-11	Sequence 11, Appl
13	8	0.9	859	US-11-188-743-16	Sequence 16, Appl
14	8	0.9	1151	US-11-128-420-10	Sequence 10, Appl
15	7	0.8	11	US-10-839-799-114	Sequence 114, App
16	7	0.8	11	US-10-839-799-129	Sequence 129, App
17	7	0.8	11	US-11-108-135-7	Sequence 7, Appli
18	7	0.8	20	US-11-040-159-4	Sequence 4, Appli
19	7	0.8	20	US-11-054-669-107	Sequence 107, App
20	7	0.8	28	US-11-174-089-152	Sequence 152, App
21	7	0.8	113	US-11-144-248-20	Sequence 20, Appl
22	7	0.8	114	US-11-065-943-47	Sequence 47, Appl
23	7	0.8	114	US-11-055-163-18	Sequence 18, Appl
24	7	0.8	116	US-11-174-186-20	Sequence 20, Appl
25	7	0.8	116	US-11-174-186-21	Sequence 21, Appl

26	7	0.8	116	7	US-11-174-186-22	Sequence 22, Appl
27	7	0.8	116	7	US-11-174-186-23	Sequence 23, Appl
28	7	0.8	116	7	US-11-174-186-24	Sequence 24, Appl
29	7	0.8	116	7	US-11-174-186-25	Sequence 25, Appl
30	7	0.8	116	7	US-11-055-163-17	Sequence 17, Appl
31	7	0.8	117	6	US-10-839-799-132	Sequence 132, App
32	7	0.8	117	6	US-10-054-669-120	Sequence 120, App
33	7	0.8	118	6	US-10-932-334-75	Sequence 75, Appl
34	7	0.8	118	7	US-11-012-353-71	Sequence 71, Appl
35	7	0.8	118	7	US-11-009-939-12	Sequence 12, Appl
36	7	0.8	120	6	US-10-932-334-71	Sequence 71, Appl
37	7	0.8	120	7	US-11-096-074-2	Sequence 2, Appli
38	7	0.8	120	7	US-11-009-939-2	Sequence 2, Appli
39	7	0.8	120	7	US-11-107-028-17	Sequence 17, Appl
40	7	0.8	121	7	US-11-108-135-24	Sequence 24, Appl
41	7	0.8	121	7	US-11-127-677-54	Sequence 54, Appl
42	7	0.8	121	7	US-11-107-028-16	Sequence 16, Appl
43	7	0.8	121	7	US-11-107-028-18	Sequence 18, Appl
44	7	0.8	121	7	US-11-107-028-19	Sequence 19, Appl
45	7	0.8	121	7	US-11-107-028-20	Sequence 20, Appl

ALIGNMENTS

RESULT 1
US-11-188-743-17
; Sequence 17, Application US/11188743
; Publication No. US20050272140A1
; GENERAL INFORMATION:
; APPLICANT: Nicolaides, Nicholas
; APPLICANT: Sasse, Philip
; APPLICANT: Kinzler, Kenneth
; APPLICANT: Grasso, Luigi
; APPLICANT: Vogelstein, Bert
; TITLE OF INVENTION: Methods for generating hypermutable
; FILE OF INVENTION: Yeast
; FILE REFERENCE: 01107.00097
; CURRENT APPLICATION NUMBER: US/11/188,743
; CURRENT FILING DATE: 2005-07-26
; PRIOR APPLICATION NUMBER: US/10/641,068
; PRIOR FILING DATE: 2003-08-15
; PRIOR APPLICATION NUMBER: US/09/788,657
; PRIOR FILING DATE: 2001-02-21
; PRIOR APPLICATION NUMBER: 60/184,336
; PRIOR FILING DATE: 2000-02-23
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 17
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-188-743-17

Query Match	100.0%;	Score 932;	DB 7;	Length 932;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 932;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	MKQLPAATVRLSSQIITSVVVKELIENSIDAGATSDVKLENYGDKIEVRDNGEG	60	
DB	1	MKQLPAATVRLSSQIITSVVVKELIENSIDAGATSDVKLENYGDKIEVRDNGEG	60	
QY	61	IKAVDAPVAMKYTSKINSHEDLENLTYYGFRGEALSGICCIAEVLITRTAADNFSTQ	120	
DB	61	IKAVDAPVAMKYTSKINSHEDLENLTYYGFRGEALSGICCIAEVLITRTAADNFSTQ	120	
QY	121	YVLDSGHILSKPRLHGGTTVTALRPFNLVPRKQFYSTAKCKDEIKKIODLLMSFG	180	
DB	121	YVLDSGHILSKPRLHGGTTVTALRPFNLVPRKQFYSTAKCKDEIKKIODLLMSFG	180	
QY	181	ILKPLDIRIVFVNKAVIWKSRVSDHKMALMSVLGTAVNMNMESFOYHSEESQIYLSGFL	240	
DB	181	ILKPLDIRIVFVNKAVIWKSRVSDHKMALMSVLGTAVNMNMESFOYHSEESQIYLSGFL	240	

```
QY 241 PKCDADHSFTSLSTPERSIFINSRPVHOKDILKIRHYNKCLKESTRLYPVFLKID 300
Db 241 PKCDADHSFTSLSTPERSIFINSRPVHOKDILKIRHYNKCLKESTRLYPVFLKID 300
QY 301 VPTADVNLTPDKSQVLLQNKESVLIALENLMTTCYGPLPSTNSYNNKTDVSAADIVL 360
Db 301 VPTADVNLTPDKSQVLLQNKESVLIALENLMTTCYGPLPSTNSYNNKTDVSAADIVL 360
QY 361 SKTAETDVLNFKVSSGKNYSNVDTSVIPFQNDMNDSEGNKTDCLNHQISIGDFGYGH 420
Db 361 SKTAETDVLNFKVSSGKNYSNVDTSVIPFQNDMNDSEGNKTDCLNHQISIGDFGYGH 420
QY 421 CSSEISNIDKNTKNAFQDISMSNVSWNSQTEYSKTCFISSVKHTQSENGKNDHIDEGE 480
Db 421 CSSEISNIDKNTKNAFQDISMSNVSWNSQTEYSKTCFISSVKHTQSENGKNDHIDEGE 480
QY 481 NEEEAAGLENSSEISADEWSRGNILKNSVGENIEPVKILVPEKSLPCKVSNNNYPPIEQWN 540
Db 481 NEEEAAGLENSSEISADEWSRGNILKNSVGENIEPVKILVPEKSLPCKVSNNNYPPIEQWN 540
QY 541 LNEDESCNKKSNVIDNKSQVLLQNKESVLIALENLMTTCYGPLPSTNSYNNKTDVSAADIVL 600
Db 541 LNEDESCNKKSNVIDNKSQVLLQNKESVLIALENLMTTCYGPLPSTNSYNNKTDVSAADIVL 600
QY 601 ATLOIEELWKTLSBEEKLYBEKATKDLERNYSOMKRAIEOSOMSLKDGRRKKIKPTSAM 660
Db 601 ATLOIEELWKTLSBEEKLYBEKATKDLERNYSOMKRAIEOSOMSLKDGRRKKIKPTSAM 660
QY 661 NLAQKHKLKTSLSNQPKDELLOQSIQEKRRSQNIKWQIIPFSMKNLKINFKKQNKVDLEE 720
Db 661 NLAQKHKLKTSLSNQPKDELLOQSIQEKRRSQNIKWQIIPFSMKNLKINFKKQNKVDLEE 720
QY 721 KDEPCLIHNLRPDPDAWMTSKTEVMLNPNRYVEEALLFKRLLENHKLPAEPLEKPIMLTE 780
Db 721 KDEPCLIHNLRPDPDAWMTSKTEVMLNPNRYVEEALLFKRLLENHKLPAEPLEKPIMLTE 780
QY 781 SLFNGSHYLDVLYKMTADDQRYSGSTYLSDDPLTANGFKIKLIPGVISITENYLEIEGMAN 840
Db 781 SLFNGSHYLDVLYKMTADDQRYSGSTYLSDDPLTANGFKIKLIPGVISITENYLEIEGMAN 840
QY 841 CLPFFGVADLKEILNAILNRNAKEVYECRPRKVISYLEGEAVRLSRQLPMYLSKEDIQDI 900
Db 841 CLPFFGVADLKEILNAILNRNAKEVYECRPRKVISYLEGEAVRLSRQLPMYLSKEDIQDI 900
QY 901 IYRMKHQFGNEIKECVHGRRPFFHLLTYLPETT 932
Db 901 IYRMKHQFGNEIKECVHGRRPFFHLLTYLPETT 932
```

```
RESULT 2
US-11-188-743-18
; Sequence 18, Application US/11188743
; Publication No. US20050272140A1
; GENERAL INFORMATION:
; APPLICANT: Nicolaides, Nicholas
; APPLICANT: Sasse, Philip
; APPLICANT: Kinzler, Kenneth
; APPLICANT: Grasso, Luigi
; APPLICANT: Vogelstein, Bert
; TITLE OF INVENTION: Methods for generating hypermutable
; TITLE OF INVENTION: Yeast
; FILE REFERENCE: 01107.00097
; CURRENT APPLICATION NUMBER: US/11/188,743
; CURRENT FILING DATE: 2005-07-26
; PRIOR APPLICATION NUMBER: US/10/641,068
; PRIOR FILING DATE: 2003-08-15
; PRIOR APPLICATION NUMBER: US/09/788,657
; PRIOR FILING DATE: 2001-02-21
; PRIOR APPLICATION NUMBER: 60/184,336
; PRIOR FILING DATE: 2000-02-23
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: fastseq for Windows Version 3.0
```

```
; SEQ ID NO 18
; LENGTH: 932
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-188-743-18
Query Match 100.0%; Score 932; DB 7; Length 932;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 932; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MKQPAATVRLSSQIITSVVVKELIENSLDAGATSDVKLENYGDFKIEVRDNGEG 60
Db 1 MKQPAATVRLSSQIITSVVVKELIENSLDAGATSDVKLENYGDFKIEVRDNGEG 60
QY 61 IKAVDAPVMAMKYTTSKINSHEDLENLTTCYGRGALGSIICIAEVLITTRTAADFSTQ 120
Db 61 IKAVDAPVMAMKYTTSKINSHEDLENLTTCYGRGALGSIICIAEVLITTRTAADFSTQ 120
QY 121 YVLDGSGHILSQKPSHLGGTITVTRALRFLKPLVRKQFYSTAKCKBIKKIODLLMSFG 180
Db 121 YVLDGSGHILSQKPSHLGGTITVTRALRFLKPLVRKQFYSTAKCKBIKKIODLLMSFG 180
QY 181 ILKPDRLRIVFVHNKAVIWKSRVSDHKALMSVLGTAVMNNMESFQYHSEESQIYLSGFL 240
Db 181 ILKPDRLRIVFVHNKAVIWKSRVSDHKALMSVLGTAVMNNMESFQYHSEESQIYLSGFL 240
QY 241 PKCDADHSFTSLSTPERSIFINSRPVHOKDILKIRHYNKCLKESTRLYPVFLKID 300
Db 241 PKCDADHSFTSLSTPERSIFINSRPVHOKDILKIRHYNKCLKESTRLYPVFLKID 300
QY 301 VPTADVNLTPDKSQVLLQNKESVLIALENLMTTCYGPLPSTNSYNNKTDVSAADIVL 360
Db 301 VPTADVNLTPDKSQVLLQNKESVLIALENLMTTCYGPLPSTNSYNNKTDVSAADIVL 360
QY 361 SKTAETDVLNFKVSSGKNYSNVDTSVIPFQNDMNDSEGNKTDCLNHQISIGDFGYGH 420
Db 361 SKTAETDVLNFKVSSGKNYSNVDTSVIPFQNDMNDSEGNKTDCLNHQISIGDFGYGH 420
QY 421 CSSEISNIDKNTKNAFQDISMSNVSWNSQTEYSKTCFISSVKHTQSENGKNDHIDEGE 480
Db 421 CSSEISNIDKNTKNAFQDISMSNVSWNSQTEYSKTCFISSVKHTQSENGKNDHIDEGE 480
QY 481 NEEEAAGLENSSEISADEWSRGNILKNSVGENIEPVKILVPEKSLPCKVSNNNYPPIEQWN 540
Db 481 NEEEAAGLENSSEISADEWSRGNILKNSVGENIEPVKILVPEKSLPCKVSNNNYPPIEQWN 540
QY 541 LNEDESCNKKSNVIDNKSQVLLQNKESVLIALENLMTTCYGPLPSTNSYNNKTDVSAADIVL 600
Db 541 LNEDESCNKKSNVIDNKSQVLLQNKESVLIALENLMTTCYGPLPSTNSYNNKTDVSAADIVL 600
QY 601 ATLOIEELWKTLSBEEKLYBEKATKDLERNYSOMKRAIEOSOMSLKDGRRKKIKPTSAM 660
Db 601 ATLOIEELWKTLSBEEKLYBEKATKDLERNYSOMKRAIEOSOMSLKDGRRKKIKPTSAM 660
QY 661 NLAQKHKLKTSLSNQPKDELLOQSIQEKRRSQNIKWQIIPFSMKNLKINFKKQNKVDLEE 720
Db 661 NLAQKHKLKTSLSNQPKDELLOQSIQEKRRSQNIKWQIIPFSMKNLKINFKKQNKVDLEE 720
QY 721 KDEPCLIHNLRPDPDAWMTSKTEVMLNPNRYVEEALLFKRLLENHKLPAEPLEKPIMLTE 780
Db 721 KDEPCLIHNLRPDPDAWMTSKTEVMLNPNRYVEEALLFKRLLENHKLPAEPLEKPIMLTE 780
QY 781 SLFNGSHYLDVLYKMTADDQRYSGSTYLSDDPLTANGFKIKLIPGVISITENYLEIEGMAN 840
Db 781 SLFNGSHYLDVLYKMTADDQRYSGSTYLSDDPLTANGFKIKLIPGVISITENYLEIEGMAN 840
QY 841 CLPFFGVADLKEILNAILNRNAKEVYECRPRKVISYLEGEAVRLSRQLPMYLSKEDIQDI 900
Db 841 CLPFFGVADLKEILNAILNRNAKEVYECRPRKVISYLEGEAVRLSRQLPMYLSKEDIQDI 900
QY 901 IYRMKHQFGNEIKECVHGRRPFFHLLTYLPETT 932
Db 901 IYRMKHQFGNEIKECVHGRRPFFHLLTYLPETT 932
```

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 26, 2005, 05:00:57 ; Search time 10058 Seconds
(without alignments)
17310.747 Million cell updates/sec

Title: US-10-079-429A-3
Perfect score: 3063
Sequence: 1 ggacagagggtgcttgcg.....aacgtaataaataataac 3063

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 5883141 seqs, 28421725653 residues
Total number of hits satisfying chosen parameters: 11766282

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*
1: gb.ba.*
2: gb.in.*
3: gb.env.*
4: gb.om.*
5: gb.ov.*
6: gb.pat.*
7: gb.ph.*
8: gb.pr.*
9: gb.ro.*
10: gb.sts.*
11: gb.sy.*
12: gb.un.*
13: gb.vi.*
14: gb.htg.*
15: gb.pl.*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3063	100.0	3063	BD181098	Human DNA
2	3063	100.0	3063	CS020326	Sequence
3	3063	100.0	3063	CS056082	Sequence
4	3063	100.0	3063	AR217665	Sequence
5	3063	100.0	3063	AR342820	Sequence
6	3063	100.0	3063	AR380890	Sequence
7	3063	100.0	3063	AR382023	Sequence
8	3063	100.0	3063	AR437149	Sequence
9	3063	100.0	3063	AR540769	Sequence
10	3063	100.0	3063	AR592873	Sequence
11	3063	100.0	3063	AX214170	Sequence
12	3063	100.0	3063	AX234586	Sequence
13	3063	100.0	3063	AX775137	Sequence
14	3063	100.0	3063	HSU13695	Human homol
15	3051.4	99.6	3239	CS130851	Sequence
16	2830	92.4	2830	BC096332	Sequence
17	2826.8	92.3	2830	BC096330	Sequence
18	2555	83.4	2682	AB102875	Sequence

19	2482	81.0	2856	6	CQ720889	Sequence
20	2239	73.1	2817	8	AY540751	Homo sapi
21	2016.6	65.8	2252	8	AY540750	Homo sapi
22	1912.4	62.4	2343	8	BC096331	Homo sapi
23	1887	61.6	3045	9	BC028939	Mus muscu
24	1856	60.6	2313	8	AB102870	Homo sapi
25	1848.2	60.3	3009	9	BC061722	Rattus no
26	1845	60.2	2004	8	AB102869	Homo sapi
27	1222.6	39.9	3147	5	AJ719999	Gallus ga
28	1160.2	37.9	1668	8	AB102872	Homo sapi
29	896.6	29.3	3276	5	BC044098	Xenopus l
30	894.2	29.2	104140	8	AC008122	Homo sapi
31	894.2	29.2	106234	8	AY267352	Homo sapi
32	891	29.1	155122	14	AC141847	Pan trogl
33	891	29.1	212131	14	AC142554	Pan trogl
34	857.4	28.0	188715	14	AC141850	Papio anu
35	849	27.7	2894	5	BC089718	Xenopus t
36	816	26.6	184340	14	AC155201	Callithri
37	701.4	22.9	705	8	AB102876	Homo sapi
38	584.8	19.1	747	8	AB102874	Homo sapi
39	583.2	19.0	591	8	AB102873	Homo sapi
40	582	19.0	588	8	AB102877	Homo sapi
41	563.6	18.4	2058	5	CR762073	Xenopus t
42	555.8	18.1	994	5	CR353667	Gallus ga
43	496	16.2	1906	8	BC008410	Homo sapi
44	492.6	16.1	1329	5	BX929478	Gallus ga
45	489.4	16.0	965	8	BC084548	Homo sapi

ALIGNMENTS

RESULT 1
BD181098 Human DNA mismatch repair proteins.
LOCUS BD181098 3063 bp DNA linear PAT 15-MAY-2003
DEFINITION Human DNA mismatch repair proteins.
ACCESSION BD181098
VERSION BD181098.1 GI:30792016
KEYWORDS JP 2002325588-A/2.
SOURCE Homo sapiens (human)

ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominiidae; Homo.

REFERENCE
1 (bases 1 to 3063)
Haseltine, W.A., Ruben, S.M., Wei, Y.F., Adams, M.D., Fleischmann, R.D.,
Frazer, C.M., Fuldner, R.A., Kirkness, E.F. and Rosen, C.A.

TITLE
Human DNA mismatch repair proteins
JOURNAL
Patent: JP 2002325588-A 2 12-NOV-2002;
HUMAN GENOME SCIENCES INC

COMMENT
OS Homo sapiens (human)
PN JP 2002325588-A/2
PD 12-NOV-2002
PF 25-JAN-2002 JP 2002016830
PR 27-JAN-1994 US 08/187757,16-MAR-1994 US 08/210143 PR
23-AUG-1994 US 08/294312

PI WILLIAM A. HASELTINE, STEVEN M. RUBEN, YING FEI WEI, MARK D. ADAMS,
PI ROBERT D. FLEISCHMANN, CLAIRE M. FRASER, REBECCA A. FULDNER, EWEN F.
PI KIRKNESS,
PI CRAIG A. ROSEN

PC C12N15/09, C07K14/47, C12P21/02, C12Q1/68// (C12P21/02, C12R1/19),
PC C12N15/00
CC Human DNA mismatch repair proteins
FH Key Location/Qualifiers
FT CDS (81).. (2879).

FEATURES
source
1..3063
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"

ORIGIN
Query Match 100.0%; Score 3063; DB 6; Length 3063;

Best Local Similarity 100.0%; Pred. No. 0;			
Matches 3063; Conservative 0; Mismatches 0; Indels 0; Gaps 0;			
QY	1	GGCAGAGTGGCTGCTTGGCGCTAGTGAGTGGTAATGGCTTCGCTCGCGCTAGCAGCAAG	60
Db	1	GGCAGAGTGGCTGCTTGGCGCTAGTGAGTGGTAATGGCTTCGCTCGCGCTAGCAGCAAG	60
QY	61	CTGCTCTGTTAAAGCGAAATGAAACAAATGGCTTCGCGCAACAGTTCGACTCCTTTCAA	120
Db	61	CTGCTCTGTTAAAGCGAAATGAAACAAATGGCTTCGCGCAACAGTTCGACTCCTTTCAA	120
QY	121	GTTCTCAGATCATCTTCGGTGGTCAAGTGTGTGTAAGAGCTTATTTGAAACCTCCTTGG	180
Db	121	GTTCTCAGATCATCTTCGGTGGTCAAGTGTGTGTAAGAGCTTATTTGAAACCTCCTTGG	180
QY	181	ATGCTGGTCCACAGCGTAGATGTTAACTGGAGAACTATGGATTTGATAAAATTCGAG	240
Db	181	ATGCTGGTCCACAGCGTAGATGTTAACTGGAGAACTATGGATTTGATAAAATTCGAG	240
QY	241	TGCGAGATAACGGGAGGGTATCAAGGCTGTGTGACCTGTAAATGGCAATGAAGTACT	300
Db	241	TGCGAGATAACGGGAGGGTATCAAGGCTGTGTGACCTGTAAATGGCAATGAAGTACT	300
QY	301	ACACCTTCAAAAATAAATAGTCAAGATCTTGAAATTTGACAACTTACGGTTTCGCTG	360
Db	301	ACACCTTCAAAAATAAATAGTCAAGATCTTGAAATTTGACAACTTACGGTTTCGCTG	360
QY	361	GAGAAGCCTTGGGCTCAATTTCTGTATAGCTGAGGTTTAAATACAAAGAACGGCTG	420
Db	361	GAGAAGCCTTGGGCTCAATTTCTGTATAGCTGAGGTTTAAATACAAAGAACGGCTG	420
QY	421	CTGATAATTTAGCACCCAGTATGTTTATAGTGGCAGTGGCCACATCTTCTCAGAAAC	480
Db	421	CTGATAATTTAGCACCCAGTATGTTTATAGTGGCAGTGGCCACATCTTCTCAGAAAC	480
QY	481	CTTCAATCTTGGTCAAGGTCAACTGTAACCTGCTTTAAGATTTATTAAGAACTACTCTG	540
Db	481	CTTCAATCTTGGTCAAGGTCAACTGTAACCTGCTTTAAGATTTATTAAGAACTACTCTG	540
QY	541	TAAGAAAGCAGTTTACTCAACTGCAGAAATAATGTAAGATGAAATTAAGAAATCCAAAG	600
Db	541	TAAGAAAGCAGTTTACTCAACTGCAGAAATAATGTAAGATGAAATTAAGAAATCCAAAG	600
QY	601	ATCTCTCATGAGCTTTGCTTAACTTAACTGACTTAAGGATTTGCTTTGTACATAACA	660
Db	601	ATCTCTCATGAGCTTTGCTTAACTTAACTGACTTAAGGATTTGCTTTGTACATAACA	660
QY	661	AGGCAGTTATTTGGCAGAAAAAGCAGATACAGATCACAAGATGGCTCTCATGTCAGTTC	720
Db	661	AGGCAGTTATTTGGCAGAAAAAGCAGATACAGATCACAAGATGGCTCTCATGTCAGTTC	720
QY	721	TGGGAGCTGCTGTTATGAAACAATATGGAATCTTTTCAGTACCACTCTGAAGAACTCAGA	780
Db	721	TGGGAGCTGCTGTTATGAAACAATATGGAATCTTTTCAGTACCACTCTGAAGAACTCAGA	780
QY	781	TTTTATCTCAGTGGATTTCTTCCAAAGTGTGATGAGACCACTCTTTCACTAGTCTTTCAA	840
Db	781	TTTTATCTCAGTGGATTTCTTCCAAAGTGTGATGAGACCACTCTTTCACTAGTCTTTCAA	840
QY	841	CACCGAAGAAAGTTCATCTTTCATAAAACAGTCGACCACTCATCAAAAAGATATCTTAA	900
Db	841	CACCGAAGAAAGTTCATCTTTCATAAAACAGTCGACCACTCATCAAAAAGATATCTTAA	900
QY	901	AGTTAATCCGACATCATTAACAATCTGAAATGGCTTAAGAACTACTCTGTTGTATCCTG	960
Db	901	AGTTAATCCGACATCATTAACAATCTGAAATGGCTTAAGAACTACTCTGTTGTATCCTG	960
QY	961	TTTTCTTTCTGAAAATCGATGTTCTTACAGCTGATGTTGATGTAATTTAAACACAGATA	1020
Db	961	TTTTCTTTCTGAAAATCGATGTTCTTACAGCTGATGTTGATGTAATTTAAACACAGATA	1020
QY	1021	AAAGCCAAAGTATTTATACAAAATAAGGAATCTGTTTAAATGCTCTTGAAAAATCTGATGA	1080
Db	1021	AAAGCCAAAGTATTTATACAAAATAAGGAATCTGTTTAAATGCTCTTGAAAAATCTGATGA	1080

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model
Run on: December 26, 2005, 00:25:02 ; Search time 1173 Seconds
(without alignments)
17403.200 Million cell updates/sec

Title: US-10-079-429A-3
Perfect score: 3063
Sequence: 1 ggcagagtggtgcttgcg.....aacgtaataaactaataac 3063

Scoring table: IDENTITY_NVC
Gapop 10.0 , Gapext 1.0

Searched: 4996997 seqs, 3332346308 residues

Total number of hits satisfying chosen parameters: 9993994

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database : N_Geneseq_21.*
- 1: Geneseqn1980s.*
 - 2: Geneseqn1990s.*
 - 3: Geneseqn2000s.*
 - 4: Geneseqn2001as.*
 - 5: Geneseqn2001bs.*
 - 6: Geneseqn2002as.*
 - 7: Geneseqn2002bs.*
 - 8: Geneseqn2003as.*
 - 9: Geneseqn2003bs.*
 - 10: Geneseqn2003cs.*
 - 11: Geneseqn2003ds.*
 - 12: Geneseqn2004as.*
 - 13: Geneseqn2004bs.*
 - 14: Geneseqn2005s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3063	100.0	3063	2	Aa97526 Human DNA
2	3063	100.0	3063	4	Aah76365 Human PMS
3	3063	100.0	3063	5	Aah75042 Nucleotid
4	3063	100.0	3063	6	Abk86089 Human PMS
5	3063	100.0	3063	6	Aad39198 Human mis
6	3063	100.0	3063	6	Aal48698 Human mis
7	3063	100.0	3063	6	Aad45354 Human MLH
8	3063	100.0	3063	6	Aad39770 Human PMS
9	3063	100.0	3063	8	Adx12940 DNA encod
10	3063	100.0	3063	8	Aca89704 cDNA enco
11	3063	100.0	3063	9	Adcl3762 Human Mut
12	3063	100.0	3063	9	Aal57764 Human mis
13	3063	100.0	3063	9	Ada06245 DNA encod
14	3063	100.0	3063	10	Adc89608 Human PMS
15	3063	100.0	3063	10	Ades5236 Farnesyl
16	3063	100.0	3063	10	Adfl17893 Human PMS
17	3063	100.0	3063	10	Adg62892 Human PMS
18	3063	100.0	3063	10	Adh62630 Human mis
19	3063	100.0	3063	10	Adh60982 Human CDN

20	3063	100.0	3063	11	ADI32109	Adi32109 Human cDN
21	3063	100.0	3063	12	ADP78841	Adf78841 Human mis
22	3063	100.0	3063	12	ADG46768	Adg46768 Human MMR
23	3063	100.0	3063	12	ADO40067	Ado40067 Human PMS
24	3063	100.0	3063	12	ADP66679	Adp66679 Human mis
25	3063	100.0	3063	13	ADRI3884	Adri3884 Human DNA
26	3063	100.0	3063	13	ADT98687	Adt98687 Human PMS
27	3063	100.0	3063	13	ADU77023	Adu77023 Human mis
28	3063	100.0	3063	13	ADX84176	Adx84176 Human lym
29	3063	100.0	3063	14	ADX58474	Adx58474 Nucleotid
30	3063	100.0	3063	14	ADY53430	Ady53430 Human PMS
31	3061.4	99.9	3063	6	AA45433	Ad45433 Human MLH
32	3051.4	99.6	3239	14	AE22842	Ae22842 Human col
33	2991	97.6	3121	14	ADX08122	Adx08122 Cyclin-de
34	2486.6	81.2	2793	13	ACN42335	Acn42335 Human dia
35	421	13.7	493	13	ADQ79340	Adq79340 Novel can
36	297.2	9.7	327	8	ABX12507	Abx12507 DNA repai
37	297.2	9.7	327	9	ADA06170	Ada06170 Human EST
38	276.8	9.0	559	5	ABV48030	Abv48030 Human pro
39	180	5.9	189	2	AAT19265	Aat19265 Human gen
40	178.4	5.8	184	3	AAC22447	Aac22447 Human sec
41	171	5.6	534	5	ABV18243	Abv18243 Human pro
42	126	4.1	478	9	ACH23591	Ach23591 Human adu
43	112.4	3.7	478	13	ADR29677	Adr29677 Mouse gen
44	101.2	3.3	2772	8	ACA89712	Aca89712 cDNA enco
45	101.2	3.3	2772	12	ADP66717	Adp66717 A. thalia

ALIGNMENTS

RESULT 1
AAQ97526
ID AAQ97526 standard; cDNA; 3063 BP.

XX
AC AAQ97526;

XX
DT 27-MAR-1996 (first entry)

XX
DE Human DNA repair protein hMLH2 coding sequence.

XX
KW DNA repair protein; hMLH1; hMLH2; hMLH3; therapy; cancer; vectors;
KW DNA synthesis; diagnosis; disease; mutL4; ds.

XX
OS Homo sapiens.

XX
FH Key
CDS Location/Qualifiers
81..2873

FT /tag= a
FT /product= "DNA repair protein."
FT /note= "The CDS may extend to position 2879 where a TGA
stop codon is located. There could possibly be an error
in the decoded protein in the specification since, if the
stop codon is the correct translation termination signal,
the C-terminal end of the protein should end with two
Threonine residues. It is possible that one of these has
been omitted."

XX
MO9520678-A1.

XX
PD 03-AUG-1995.

XX
PP 25-JAN-1995; 95WO-US001035.

XX
PR 27-JAN-1994; 94US-00187757.

XX
PR 16-MAR-1994; 94US-00210143.

XX
PR 23-AUG-1994; 94US-00294312.

XX
PA (HUMA-) HUMAN GENOME SCI INC.

XX
PI Haseltine WA, Ruben SM, Wei Y, Adams MD, Fleischmann RD;
PI Fraser CM, Fuldner RA, Kirkness EF, Rosen CA;

WPI; 1995-275461/36.
P-PSDB; AAR79009.

P-PSDB; AAR79009.

Polynucleotide(s) encoding human mutL homologues, hMLH1, hMLH2 and hMLH3 - used for therapeutic treatment of, e.g. hereditary cancer.

- used for therapeutic treatment of, e.g. hereditary cancer.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

Claim 1. Fig 2: 124nn. English.

12/21/2004

The polynucleotides described in AAQ97525-27 encode the human analogues of the prokaryotic mutL DNA repair gene. The polypeptides they encode (AAQ97525-27) are used for therapeutic purposes e.g. in the treatment of cancer, esp. hereditary cancer. They may also be used for in vitro manipulation of DNA, synthesis of DNA and the manufacture of DNA vectors and in methods of diagnosing a disease or a susceptibility to a disease related to a mutation in the bMLH1 -2 or -3 DNA repair genes.

related to a mutation in the *hprt* gene.

COMM-2022-3067 BP: 1100 A: 503 C: 580 G: 880 T: 0 U: 0 Other:

[illegible]

Query Match	Score	Score
100.00	3063	100.00

Best Local Similarity	100.0%; Pred. No. 0;	100.0%; Pred. No. 0;	100.0%; Pred. No. 0;

5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843

1 GGCACGAGTGGCTGCTTGGGCTAGTGGAATGCTGCTCGGCCTAGCAGCAAG 60

1 GGCACGAGTGGCTGCTTGGGGCTAGTGGATGGTAATTGCCCTCGCGCTAGCAGCAAG 60

61 CTGGCTCTGTTAAAGCGAAAATGAAACAATTGCCTGCGGCAACAGTTCGACTCTCTTTTCAA 120

101 CATTCCGACATCATCACTTCAGGTCTCAGTGTCTATAAAGAGGCTTATTTGAAACTCCTTGG 180

121 G T C A G A T C A C A T C C G G T C A G T G T G A T F R S N C

121 GTCACAGATCATCATTCTGGGTCAGTGTTAAAGAAGCTAATCAATTAATCCCTTCT

.....

1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525

181 ATGCTGGTGCCCAAGCGTAGATGTTAAAC TGGAGAAC TATGGATTTGATAAAAATGAGG 240

241 TCGAGATAACGGGGAGGGTATCAAGGCTGTTGATGCACCTGTAAATGGCAATGAAGTACT 300

.....

301 ACACCTCAAAAATAAATAGTCATGAAGATCTTGAAATTTGACAACCTTACGGTTTTCGTG 360

301 ACCCTCAAAATATATCTCATCAAGATCTTGACAACTTACGGTTTCGTG 360

301 ACACCCCTCATTATTAATAAGTCATGCAAAATCTTGTGGTTAATGCGCATG

361 GAGAAGCTTGGGTCAAATTTGTGTGTAAGTAGGTTTTAATTACCAACAAGAACCGCCCTC

[illegible]

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

421 CTGATAATTTAGCACCCAGTATGTTTAGATGGCAGTGGCCACATACITTTCTCAGAAAC 480

421 CTGATAATTTAGCACCCAGTATGTTTATAGTGGCAGTGGCCACATACCTTCTCAGAAAC 480

481 CTTTCA CATCTTTGGTCAAGGTCAACTGTAACTGCTTTAAGATTATTTAAGAACTACCTG 540

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 26, 2005, 05:14:07 ; Search time 7732 Seconds
(without alignments)
18534.510 Million cell updates/sec

Title: US-10-079-429A-3
Perfect score: 3063
Sequence: 1 ggcagagtggctgtctgg.....aacgtataataactaataac 3063

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 41078325 seqs, 23393541228 residues

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*
10: gb_est10:*
11: gb_est11:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2969.8	97.0	3130	CR859593	CR859593 Pongo pyg
2	2164.2	70.7	2365	BC036376	BC036376 Homo sapi
3	1970	64.3	2667	DQ052952	DQ052952 Homo sapi
4	1825.8	59.6	2667	DQ052953	DQ052953 Pan trogl
5	1804	58.9	1804	CR610658	CR610658 full-length
6	891.2	29.1	1011	EX353664	EX353664 BX353664
7	846	27.6	993	CO580445	CO580445 ILLUMIN
8	835.4	27.3	893	EX328949	EX328949 BX328949
9	821.8	26.8	903	EX327629	EX327629 BX327629
10	819.2	26.7	1021	BM479838	BM479838 AGENCOURT
11	803.6	26.2	898	EX435290	EX435290 BX435290
12	780.6	25.5	1117	EM553209	EM553209 AGENCOURT
13	780	25.5	780	CN336501	CN336501 170006001
14	776	25.3	776	EX117693	EX117693 BX117693
15	721	23.5	733	CN336498	CN336498 170006001
16	718	23.4	851	DN102367	DN102367 1097287 M
17	717.8	23.4	857	AM019977	AM019977 AM019977
18	708.2	23.1	820	EX419980	EX419980 BX419980
19	706.2	23.1	801	EG193431	EG193431 RST12563
20	704.6	23.0	986	EX353663	EX353663 BX353663
21	702.8	22.9	842	DN526212	DN526212 1270959 M
22	700.2	22.9	956	BM800196	BM800196 AGENCOURT

C 23	694	22.7	722	3	BM677668	BM677668 UI-E-E01-
C 24	677	22.1	704	5	BQ771615	BQ771615 UI-H-EZ1-
C 25	676.6	22.1	902	8	DN523518	DN523518 1267023 M
26	650.2	21.2	694	1	AL705101	KXFZ9686H
27	640	20.9	641	6	CB157373	CB157373 K-EST0216
C 28	639.4	20.9	748	6	CA415469	CA415469 UI-H-EZ0-
C 29	637.6	20.8	706	6	CB852984	CB852984 UI-CF-FN0
C 30	634.4	20.7	746	5	BU623174	BU623174 UI-H-FL1-
C 31	634	20.7	638	1	AL043809	AL043809 DXFZP434P
32	633.2	20.7	835	2	BG168340	BG168340 602342315
33	627	20.5	753	5	EX925658	EX925658 BX925658
34	626.4	20.5	668	8	DN998411	DN998411 TC108042
35	621	20.3	843	3	BI464618	BI464618 603203232
36	620.8	20.3	782	5	BQ429685	BQ429685 AGENCOURT
37	617.2	20.2	949	5	BX452128	BX452128 BX452128
38	616.4	20.1	852	2	BG536475	BG536475 602564576
39	613.8	20.0	825	3	BI545790	BI545790 603187943
40	603.8	19.7	735	5	EX925660	EX925660 BX925660
41	602	19.7	603	3	BM723144	BM723144 UI-E-E01-
42	601.8	19.6	869	7	C0737632	C0737632 SILH03C18
43	601.8	19.6	988	2	BG163660	BG163660 602338840
44	598.2	19.5	905	2	BF666456	BF666456 602123905
C 45	596	19.5	623	2	BE350913	BE350913 ht63h02.x

ALIGNMENTS

RESULT 1
CR859593
LOCUS
DEFINITION
Pongo pygmaeus mRNA; cDNA DKFZp468M105 (from clone DKFZp468M105).
ACCESSION
CR859593.1
VERSION
CR859593.1
KEYWORDS
HTC.
SOURCE
Pongo pygmaeus (orangutan)
ORGANISM
Pongo pygmaeus
REFERENCE
1 (bases 1 to 3130)
AUTHORS
Ansoberg, W., Krieger, S., Regiert, T., Rittmueller, C., Schwager, B., Mewes, H.W., Weil, B., Amid, C., Osanger, A., Fobo, G., Han, M. and Wiemann, S.
CONSRMT
The German cDNA Consortium
TITLE
Direct Submission
JOURNAL
Submitted (12-NOV-2004) MIPS, Ingolstaedter Landstr.1, D-85764 Neuherberg, GERMANY
COMMENT
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de; sequenced by EMBL (European Molecular Biology Laboratories, Heidelberg/Germany) within the cDNA sequencing consortium of the German Genome Project.
This clone (DKFZp468M105) is available at the RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH in Berlin, Germany.
Please contact RZPD for ordering:
http://www.rzpd.de/cgi-bin/products/cl.cgi?cloneID=DKFZp468M105
Further information about the clone and the sequencing project is available at http://mips.gsf.de/projects/cdna/.
FEATURES
Location/Qualifiers
1..3130
/organism="Pongo pygmaeus"
/mol_type="mRNA"
/db_xref="taxon:9600"
/clone="DKFZp468M105"
/tissue_type="heart"
/clone_lib="468 (synonym: phrt1). Vector pSport1_Sfi; host DH10B; sites SfiI + SfiII"
/dev_stage="adult"
/note="PMS1 protein homolog 1 (Homo sapiens)"
1..3130
/gene="DKFZp468M105"
135..2930
CDS

```

/gene="DKFZp468M105"
/codon_start=1
/product="hypothetical protein"
/protein_id="CAH91756.1"
/db_xref="GI:55730063"
/translation="MKQLPAAVRLSSQIITSVVVVKELIENSLDAGATSIDVKL
ENYGDPIKLEVRDNGSGIKAVDTPVMAMKYTSKINSHEDLENLTYYGFRGEALGICC
VAEVLITTRTAADNSTQVYLDGSHIILSQKPSHLAQGTVTVALRLPRLPVKQFYS
TAKKDEIKIQQDLMSYGLIKPDLIRIVFHNKAVIWKSRVSHKALMSVLGTAV
MSNMESFQHSSESQIYLSGFLPKDADHSFTSLSTPERSFIFINSRPVHQKIDILKI
RHHYVLKCLKESTRIPYVFLKIDVPTADVDNLTPDKSQVLLQNKESILALELMT
TCVGPLLSYNNKNTDVSAAIDVLSKTAETDVLFNKMSSESKNYSNVDTSIIPFON
DHNDESIGNKNDLCHOLISIGDFGVGHCSSEISNVDKNTKNAFODISMNSVWNTQ
TEYSKTCFVGSIKHTQSENGKNDHIDESGENEERAGLENSLEISADENSRGNILNSV
GENIEBPKLVPEKSLPCVSNNNYPSPEQTINLEDPCNKNKNVINDSKSKYKATYDLL
SNRVKPMASALFQDRHPQLIENPTSLIEDATLQIEELWKTLSSEBKUYEKA
TKDLERYNNQMRATEQESQISLKDGKKIKPTSAWNLAKRHKJLTSLSNPKLDLKL
QSOIERSONIKWQVIFPMKNLKNFKKQDLEEKNEPCLIHNLRFPPDAMLWTSK
TEVMLNLPYRVEALLFKELLENHKLPAPELEKPIMLTLESLFNGSHYLDVLYKMTADD
QYSGSTYLSDEPLRTANGPKIKLIPQVSITENYLEEGMANCLPFVGVADLKEIILNAI
LNKNAKEVTECPRKVISYLEGEAVRLSKQLPWLKSDIQDIIIRMKHQFQNEIKGC
VHGRPFPHLTYLPETT"

```

ORIGIN

Query Match		97.0%;	Score 2969.8;	DB 4;	Length 3130;
Best Local Similarity		98.5%;	Pred. No. 0;		
Matches 3008;		Conservative	0;	Mismatches	42;
				Indels	3;
				Gaps	1;
QY	9	TGGCTGCTTGGCGCTAGTGGATGTAATTGCGCTCGCTCGCTAGCAGCAAGCTGCTCTG	68		
DB	63	TGGCTGCTTGGCGCTAGTGGATGTAATTGCTGCTCGCTCGCTAGCAAGCTGCTCTG	122		
QY	69	TTAAAGCGAAATGAAACAAATTGCGCTGGCGCAACAGTTCGACTCCTTTCAAGTTCCTCAG	128		
DB	123	TTAAAGCGAAATGAAACAAATTGCGCTGGCGCAACAGTTCGACTCCTTTCAAGTTCCTCAG	182		
QY	129	ATCATCAGTTCGGTGTGAGTGTGTAAGAGAGCTTATTGAAACCTCTTGGATGCTGGT	188		
DB	183	ATCATCAGTTCGGTGTGAGTGTGTAAGAGAGCTTATTGAAACCTCTTGGATGCTGGT	242		
QY	189	GCCACAAGCGTAGATTTAACTGGAGAACTATGGATTTTGATAAAATTCAGGTGGCAGAT	248		
DB	243	GCCACAAGCATAGATCTTAACTGGAGAACTATGGATTTTGATAAAATTCAGGTGGCAGAT	302		
QY	249	AACGGGAGGATCAAGGCTGTGATGACACCTGTAAATGGCAATGAAGTACTACACCTCA	308		
DB	303	AACGGGAGGATCAAGGCTGTGATGACACCTGTAAATGGCAATGAAGTACTACACCTCA	362		
QY	309	AAATAAATAGTCATCAAGACTTTGAAATTTTCACAACTTACCGTTTTCTGGAGAAAGCC	368		
DB	363	AAATAAATAGTCATCAAGACTTTGAAATTTTCACAACTTACCGTTTTCTGGAGAAAGCC	422		
QY	369	TTGGGCTCAATTTGTTGTAGTGGAGTTTAAATTAACAAGAACCGGCTGCTGATAAT	428		
DB	423	TTGGGCTCAATTTGTTGTAGTGGAGTTTAAATTAACAAGAACCGGCTGCTGATAAT	482		
QY	429	TTTAGCACCAGPATGTTTGTAGTGGAGTGGCCACATCTTCTCAGAAACCTTCACAT	488		
DB	483	TTTAGCACCAGPATGTTTGTAGTGGAGTGGCCACATCTTCTCAGAAACCTTCACAT	542		
QY	489	CTTGGTCAAGGTACAACTGTAACTGCTTTTAAAGATTATTAAAGAACTTACCTGTGAAGAAAG	548		
DB	543	CTTGGTCAAGGTACAACTGTAACTGCTTTTAAAGATTATTAAAGAACTTACCTGTGAAGAAAG	602		
QY	549	CAGTTTTACTCAACTGCAAAAAAATGTAAAGATGAAATTAATAAAGATCCAAAGATCTCCTC	608		
DB	603	CAGTTTTACTCAACTGCAAAAAAATGTAAAGATGAAATTAATAAAGATCCAAAGATCTCCTC	662		
QY	609	ATGAGCTTTGGTATCCTTAACTGACTTTAAGGATTTGCTTTGTATCAATAAAGGAGATT	668		
DB	663	ATGAGCTATGGTATCCTTAACTGACTTTAAGGATTTGCTTTGTATCAATAAAGGAGATT	722		
QY	669	ATTGGCAGAAAGCAGAGTATCAGATCAAGATGGCTCTCATGTGTCAGTTCTGGGGACT	728		

DB	723	ATTGGCAGAAAGCAGAGTATCAGATCAAGATGGCTCTCATGTGCTGAGTTCTGGGGACT	782		
QY	729	GCTGTTATGAACAATATGGAATCCTTTTCAGTACCACTCTGGAAGATCTCAGATTTATCTC	788		
DB	783	GCTGTTATGAGCAATATGGAATCCTTTTCAGTACCACTCTGGAAGATCTCAGATTTATCTC	842		
QY	789	AGTGATTTCTTCCAAAGTGTGATGACACCACTCTTTTCACCTAGTCTTTCAACACAGAA	848		
DB	843	AGTGATTTCTTCCAAAGTGTGATGACACCACTCTTTTCACCTAGTCTTTCAACACAGAA	902		
QY	849	AGAAAGTTTCATCTTCATAAAACAGTCGACAGTACATCAAAAAGATATCTTAAAGTTAATC	908		
DB	903	AGAAAGTTTCATCTTCATAAAACAGTCGACAGTACATCAAAAAGATATCTTAAAGTTAATC	962		
QY	909	CGACATCATTAACAATCTGAAATGCTTAAAGGAATCTACTCGTTTCTGATCTCTTTCTTT	968		
DB	963	CGACATCATTAACAATCTGAAATGCTTAAAGGAATCTACTCGTTTCTGATCTCTTTCTTT	1022		
QY	969	CTGAAAATCGATGTTCTTACAGCTGATGTTGATGTAATTTTAAACACAGATAAAAGCCAA	1028		
DB	1023	CTGAAAATCGATGTTCTTACAGCTGATGTTGATGTAATTTTAAACACAGATAAAAGTCAA	1082		
QY	1029	GTATTAATTAACAAATAGGAATCTGTTTAAATTTGCTCTTGAATAATCTGATGACGACTGT	1088		
DB	1083	GTATTAATTAACAAATAGGAATCTGTTTAAATTTGCTCTTGAATAATCTGATGACGACTGT	1142		
QY	1089	TATGACCATTTACCTAGTACAAATCTTATGAAAATAATAAAACAGATGTTTCCGCGAGCT	1148		
DB	1143	TATGACCATTTACTTTAGTACAAATCTTATGAAAATAATAAAACAGATGTTTCCGCGAGCT	1202		
QY	1149	GACATCGTTCTTTAGTAAACACAGACAGAAACAGATGCTGTTTTTAAATAAGTGAATCATCT	1208		
DB	1203	GACATCGTTCTTTAGTAAACACAGACAGAAACAGATGCTGTTTTTAAATAAGTGAATCATCT	1262		
QY	1209	GGAAAGAAATTAATCAAAATGTTGATCTTTCAGTCAATTCATTTCCAAAATGATATGATAT	1268		
DB	1263	GGAAAGAAATTAATCAAAATGTTGATCTTTCAGTCAATTCATTTCCAAAATGATATGATAT	1322		
QY	1269	GATGAATCTGGAAAAAACAACACTGATGATTTGTTTAAATCACCAGATAAGTATTCGTGACTTT	1328		
DB	1323	GATGAATCTGGAAAAAACAACACTGATGATTTGTTTAAATCACCAGATAAGTATTCGTGACTTT	1382		
QY	1329	GCTTATGCTCAATTTAGTAGTGAAATTTCTAACTTGAATAAAAAACATTAAGATGCTATTT	1388		
DB	1383	GCTTATGCTCAATTTAGTAGTGAAATTTCTAACTTGAATAAAAAACATTAAGATGCTATTT	1442		
QY	1389	CAGGACATTTCAATGAGTAATGATATCATGGGAGAACTCTCAGACGGAATATAGTAAACT	1448		
DB	1443	CAGGACATTTCAATGAGTAATGATATCATGGGAGAACTCTCAGACGGAATATAGTAAACT	1502		
QY	1449	TGTTTTTATAAGTTCCGTTTAAAGCACACCCAGTCAGAAAAATGGCAATAAAGACCATATAGAT	1508		
DB	1503	TGTTTTTATAAGTTCCGTTTAAAGCACACCCAGTCAGAAAAATGGCAATAAAGACCATATAGAT	1562		
QY	1509	GAGAGTGGGGAAAAATGAGGAAGAAGCAGGTCTTTGAAAACTCTTCGGAATTTCTCGAGAT	1568		
DB	1563	GAGAGTGGGGAAAAATGAGGAAGAAGCAGGTCTTTGAAAACTCTTCGGAATTTCTCGAGAT	1622		
QY	1569	GAGTGGAGCAGGGGAAAAATATCTTAAAAATTCAGTGGGAGAGAAATTTGAACTGTGAAA	1628		
DB	1623	GAGTGGAGCAGGGGAAAAATATCTTAAAAATTCAGTGGGAGAGAAATTTGAACTGTGAAA	1682		
QY	1629	ATTTTATGTCCTGAAAAAAGTTTACCATGTAAGTAAGTAATAAATTAATTTTCAATCCCT	1688		
DB	1683	ATTTTATGTCCTGAAAAAAGTTTACCATGTAAGTAAGTAATAAATTAATTTTCAATCCCT	1742		
QY	1689	GAAACAAATGAATCTTAAATGAAGATTAATGTAACAAAAATCAAAATGTAATAGATAATAA	1748		
DB	1743	GAAACAAATGAATCTTAAATGAAGATTAATGTAACAAAAATCAAAATGTAATAGATAATAA	1802		
QY	1749	TCGTGAAAAAGTTTACAGCTTATGATTTTATAGCAATTCAGTAAATCAAGAAACCCATGTCA	1808		

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 26, 2005, 06:20:02 ; Search time 371 Seconds
(without alignments)
14675.664 Million cell updates/sec

Title: US-10-079-429A-3
Perfect score: 3063
Sequence: 1 ggcagagggctgctgcg.....aacgtaataaactaataac 3063

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0
Searched: 1303057 seqs, 888780828 residues
Total number of hits satisfying chosen parameters: 2606114

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA:*
1: /cgn2_6/prodata/1/ina/1 COMB.seq.*
2: /cgn2_6/prodata/1/ina/5 COMB.seq.*
3: /cgn2_6/prodata/1/ina/6A COMB.seq.*
4: /cgn2_6/prodata/1/ina/6B COMB.seq.*
5: /cgn2_6/prodata/1/ina/H COMB.seq.*
6: /cgn2_6/prodata/1/ina/PCTUS COMB.seq.*
7: /cgn2_6/prodata/1/ina/PP COMB.seq.*
8: /cgn2_6/prodata/1/ina/BE COMB.seq.*
9: /cgn2_6/prodata/1/ina/backfileseq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3063	100.0	3063	US-08-294-312B-3	Sequence 3, Appli
2	3063	100.0	3063	US-08-468-024B-3	Sequence 3, Appli
3	3063	100.0	3063	US-09-708-200-12	Sequence 12, Appl
4	3063	100.0	3063	US-09-023-655-1435	Sequence 1435, Ap
5	3063	100.0	3063	US-08-465-679-3	Sequence 3, Appli
6	3063	100.0	3063	US-09-788-657-8	Sequence 8, Appli
7	3063	100.0	3063	US-09-712-691-10	Sequence 10, Appl
8	3063	100.0	3063	US-09-707-468C-10	Sequence 10, Appl
9	3063	100.0	3063	US-10-641-068-8	Sequence 8, Appli
10	2808.4	91.7	2817	US-09-949-016-1916	Sequence 1916, Ap
11	894.2	29.2	96845	US-09-949-016-13658	Sequence 13658, A
12	297.2	9.7	327	US-08-187-757D-3	Sequence 3, Appli
13	297.2	9.7	327	US-08-210-143C-3	Sequence 3, Appli
14	178.4	5.8	184	US-09-513-999C-26522	Sequence 26522, A
15	134	4.4	601	US-09-949-016-65618	Sequence 65618, A
16	99.8	3.3	2589	US-09-749-601A-3	Sequence 3, Appli
17	99.8	3.3	2687	US-08-209-521-22	Sequence 22, Appl
18	99.8	3.3	2687	US-08-961-810-132	Sequence 132, App
19	99.8	3.3	2687	US-08-352-902D-132	Sequence 132, App
20	99.8	3.3	2687	US-09-265-503B-132	Sequence 132, App
21	99.8	3.3	2771	US-09-059-461-1	Sequence 1, Appli
22	99.8	3.3	2771	US-08-294-312B-5	Sequence 5, Appli
23	99.8	3.3	2771	US-08-468-024B-5	Sequence 5, Appli
24	99.8	3.3	2771	US-09-708-200-10	Sequence 10, Appl

ALIGNMENTS

RESULT 1

```
US-08-294-312B-3
; Sequence 3, Application US/08294312B
; Patent No. 6380369
; GENERAL INFORMATION:
; APPLICANT: Adams et al.
; TITLE OF INVENTION: Human DNA Mismatch Repair Proteins
; FILE REFERENCE: PF106P2
; CURRENT APPLICATION NUMBER: US/08/294,312B
; PRIOR FILING DATE: 1994-08-23
; PRIOR APPLICATION NUMBER: 08/210,143
; PRIOR FILING DATE: 1994-03-16
; PRIOR APPLICATION NUMBER: 08/187,757
; PRIOR FILING DATE: 1994-01-27
; NUMBER OF SEQ ID NOS: 78
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 3
; LENGTH: 3063
; TYPE: DNA
; ORGANISM: homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (81)..(2879)
US-08-294-312B-3

Query Match      100.0%; Score 3063; DB 3; Length 3063;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 3063; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGCACGAGTGGCTGCTTGGCGGTCTAGTGGATGTAATTGCTGCGCTAGCAGCAAG 60
    |||||
Db 1 GGCACGAGTGGCTGCTTGGCGGTCTAGTGGATGTAATTGCTGCGCTAGCAGCAAG 60
    |||||

QY 61 CTGCTCTGTAAAAGCGAAATGAAACAATTCCTCGCGCAACAGTTCGACTCCTTTCAA 120
    |||||
Db 61 CTGCTCTGTAAAAGCGAAATGAAACAATTCCTCGCGCAACAGTTCGACTCCTTTCAA 120
    |||||

QY 121 GTTCTCAGATCATCTTCGTTGGTGGTGTGTAAGAGAGCTTATGAAACTCTCTTGG 180
    |||||
Db 121 GTTCTCAGATCATCTTCGTTGGTGGTGTGTAAGAGAGCTTATGAAACTCTCTTGG 180
    |||||

QY 181 ATGCTGGTGCCACAGCGTAGATGTTAACTGAGAACTATGATGTTTGAATAAATTGAGG 240
    |||||
Db 181 ATGCTGGTGCCACAGCGTAGATGTTAACTGAGAACTATGATGTTTGAATAAATTGAGG 240
    |||||

QY 241 TGCAGATAACGGGGAGGGTATCAAGGGCTGTTGATGCACCTGTTGTAATGCAATGAAGTACT 300
    |||||
```

Db 241 TGGAGATTAACGGGGAGGGTATCAAGGCTGTGATGCACCTGTAAATGGCAATGAAGTACT 300
QY
301 ACACCTCAAAAAATAATAGTCATGAAGATCTTGAAAAATTTGACAACTTACGGTTTTTCGTG 360
Db 301 ACACCTCAAAAAATAATAGTCATGAAGATCTTGAAAAATTTGACAACTTACGGTTTTTCGTG 360
QY 361 GAGAGCCCTGGGGTCAATTTGTTGTTATAGTCAGGTTTTTAATTTACAAAGAAGACGGCTG 420
Db 361 GAGAGCCCTGGGGTCAATTTGTTGTTATAGTCAGGTTTTTAATTTACAAAGAAGACGGCTG 420
QY 421 CTGATAATTTTAGCACCCAGTATGTTTTAGATGGCAGTGGCCACATATCTTTCTCAGAAAC 480
Db 421 CTGATTAATTTAGCACCCAGTATGTTTTAGATGGCAGTGGCCACATATCTTTCTCAGAAAC 480
QY 481 CTTTCAATCTTGTCAGAGTAACTGTAACCTGTTAACTGTTTAAGATTTATTAAGAATCTTACCTG 540
Db 481 CTTTCAATCTTGTCAGAGTAACTGTAACCTGTTAACTGTTTAAGATTTATTAAGAATCTTACCTG 540
QY 541 TAAGAAAGCAGTTTTTACTCAACTGCAAAAAAATGTAAGATGAATAAAAAAGATCCAAG 600
Db 541 TAAGAAAGCAGTTTTTACTCAACTGCAAAAAAATGTAAGATGAATAAAAAAGATCCAAG 600
QY 601 ATCTCTCATGAGCTTTGGTATCCTTAAACCTGACTTAAGGATTTGTTTGTACATAACA 660
Db 601 ATCTCTCATGAGCTTTGGTATCCTTAAACCTGACTTAAGGATTTGTTTGTACATAACA 660
QY 661 AGGCAGTTATTTGGCAGAAAAAGCAGATATCAGATCAAGAATGGCTCTCATGTCAGTTTC 720
Db 661 AGGCAGTTATTTGGCAGAAAAAGCAGATATCAGATCAAGAATGGCTCTCATGTCAGTTTC 720
QY 721 TGGGGACTGCTGTTATGAACAAATATGAATCCTTTTCAAGTACCCTCTGAAGAATCTCAGA 780
Db 721 TGGGGACTGCTGTTATGAACAAATATGAATCCTTTTCAAGTACCCTCTGAAGAATCTCAGA 780
QY 781 TTTTATCTCAGTGGATTTCTTCCAAAGTGTGATGCAGACCACTCTTTTCACTAGTCTTTCAA 840
Db 781 TTTTATCTCAGTGGATTTCTTCCAAAGTGTGATGCAGACCACTCTTTTCACTAGTCTTTCAA 840
QY 841 CACACAGAAAGATTTTCACTTCTCAATAAAGTCGACAGTACATCAAAAAAGATATCTTAA 900
Db 841 CACACAGAAAGATTTTCACTTCTCAATAAAGTCGACAGTACATCAAAAAAGATATCTTAA 900
QY 901 AGTTATCCGACATCATTTACAACTGAAATGCCTTAAAGGAATCTACTCGTTTGTATCTCTG 960
Db 901 AGTTATCCGACATCATTTACAACTGAAATGCCTTAAAGGAATCTACTCGTTTGTATCTCTG 960
QY 961 TTTTCTTTCTGAAAAATCGATGTTCTTACAGCTGATGTTGATGTAATTTAAACACAGATA 1020
Db 961 TTTTCTTTCTGAAAAATCGATGTTCTTACAGCTGATGTTGATGTAATTTAAACACAGATA 1020
QY 1021 AAAGCCAAAGTATTTATTAACAAATTAAGGAATCTGTTTAAATGCTCTTTGAAAAATCTGATGA 1080
Db 1021 AAAGCCAAAGTATTTATTAACAAATTAAGGAATCTGTTTAAATGCTCTTTGAAAAATCTGATGA 1080
QY 1081 CGACTGTTATGGACCAATACCTAGTACAAATCTTATGAAATTAATAAACAAGATGTTT 1140
Db 1081 CGACTGTTATGGACCAATACCTAGTACAAATCTTATGAAATTAATAAACAAGATGTTT 1140
QY 1141 CCGCAGCTGACATCGTTCTTTAGTAAAAACAGCAGAAACAGATGTCGTTTTTAAATAAGTGG 1200
Db 1141 CCGCAGCTGACATCGTTCTTTAGTAAAAACAGCAGAAACAGATGTCGTTTTTAAATAAGTGG 1200
QY 1201 AATCATCTGGAAGAATTTATCAAAATGTTGATATCTTCAGTCACTTCGATTTCCAAATGATA 1260
Db 1201 AATCATCTGGAAGAATTTATCAAAATGTTGATATCTTCAGTCACTTCGATTTCCAAATGATA 1260
QY 1261 TGCATAATGATGATCTGGAAAAACACATGATGATGTTTAAATCACCAGATAGTATTG 1320
Db 1261 TGCATAATGATGATCTGGAAAAACACATGATGATGTTTAAATCACCAGATAGTATTG 1320
QY 1321 GTGACTTTGGTTATGGTCAATTGTAGTAGTGAATTTCTTAACATTTGATAAAAAACACTAAGA 1380
Db 1321 GTGACTTTGGTTATGGTCAATTGTAGTAGTGAATTTCTTAACATTTGATAAAAAACACTAAGA 1380

QY 1381 ATGCATTTTCAGGACATTTCAATGAGTAAATGTATCATGGAGAACTCTCAGACGGAATATA 1440
Db 1381 ATGCATTTTCAGGACATTTCAATGAGTAAATGTATCATGGAGAACTCTCAGACGGAATATA 1440
QY 1441 GTAAAACTTGTTTTTAATAAGTCCGTTAAGCACACCCAGTCAAGAAATGGCAATAAAGACC 1500
Db 1441 GTAAAACTTGTTTTTAATAAGTCCGTTAAGCACACCCAGTCAAGAAATGGCAATAAAGACC 1500
QY 1501 ATATAGATGAGAGTCGGGAAATATGAGGAAGAGAGAGTCTTGAAACCTCTTCGGAATTTT 1560
Db 1501 ATATAGATGAGAGTCGGGAAATATGAGGAAGAGAGAGTCTTGAAACCTCTTCGGAATTTT 1560
QY 1561 CTGCAGATGAGTGGAGCAGGGGAAATATATCTTTAAAAATTTCACTGGGAGAGAAATATTGAAC 1620
Db 1561 CTGCAGATGAGTGGAGCAGGGGAAATATATCTTTAAAAATTTCACTGGGAGAGAAATATTGAAC 1620
QY 1621 CTGTGAAAAATTTTGTGCTGAAAAAGTTTACCATGTAAAGTAAAGTAAATTAATTAATTC 1680
Db 1621 CTGTGAAAAATTTTGTGCTGAAAAAGTTTACCATGTAAAGTAAAGTAAATTAATTAATTC 1680
QY 1681 CAATCCCTGAAACAAATGAATCTTAAATGAAGTTCATGTAAACAAAAATCAAATGTAAATAG 1740
Db 1681 CAATCCCTGAAACAAATGAATCTTAAATGAAGTTCATGTAAACAAAAATCAAATGTAAATAG 1740
QY 1741 ATAAATAAATCTGAAAAAGTTTACAGCTTTATGATTTTACTTAGCAATCGAGTAAATCAAGAAC 1800
Db 1741 ATAAATAAATCTGAAAAAGTTTACAGCTTTATGATTTTACTTAGCAATCGAGTAAATCAAGAAC 1800
QY 1801 CCATGTCAGCAAGTGTCTTTTTTTTGTTCAGAGTATCTGCTCCTCAGTTTCTCATAGAAAAATC 1860
Db 1801 CCATGTCAGCAAGTGTCTTTTTTTTGTTCAGAGTATCTGCTCCTCAGTTTCTCATAGAAAAATC 1860
QY 1861 CTAAGACTAGTTTACAGGATGCAACACTACAAATTTGAAGAACTGTGGAAGACATTCAGTGTG 1920
Db 1861 CTAAGACTAGTTTACAGGATGCAACACTACAAATTTGAAGAACTGTGGAAGACATTCAGTGTG 1920
QY 1921 AAGAGGAAAACTGAAAAATATGAAGAGAGAGGCTACTAAAGACTTTGGAACGATACAAATAGTC 1980
Db 1921 AAGAGGAAAACTGAAAAATATGAAGAGAGAGGCTACTAAAGACTTTGGAACGATACAAATAGTC 1980
QY 1981 AAATGAAGAGAGCCATTGAAACAGGAGTCAAAATGTCTCAATAAGTGGCAGAAAAAAGA 2040
Db 1981 AAATGAAGAGAGCCATTGAAACAGGAGTCAAAATGTCTCAATAAGTGGCAGAAAAAAGA 2040
QY 2041 TAAACCCACACCGCATGGAATTTGGCCCGAGAGCACAAGTTAAAGAACTCATTTATCTA 2100
Db 2041 TAAACCCACACCGCATGGAATTTGGCCCGAGAGCACAAGTTAAAGAACTCATTTATCTA 2100
QY 2101 ATCAACCAAAAATCTTGATGAACCTCTTTCAGTCCCAAAATTTGAAAAAAGAGAGTCAAAATA 2160
Db 2101 ATCAACCAAAAATCTTGATGAACCTCTTTCAGTCCCAAAATTTGAAAAAAGAGAGTCAAAATA 2160
QY 2161 TTAATAATGTTACAGATCCCTCTTTTCTATGAAAAAATTAAAAATTAATTTTAAGAAAAAAA 2220
Db 2161 TTAATAATGTTACAGATCCCTCTTTTCTATGAAAAAATTAAAAATTAATTTTAAGAAAAAAA 2220
QY 2221 ACAGGTTGACTTAGAAGAGAGAGGATGAACCTTGTGATCCCAATCTCAGGTTTCCCTG 2280
Db 2221 ACAGGTTGACTTAGAAGAGAGAGGATGAACCTTGTGATCCCAATCTCAGGTTTCCCTG 2280
QY 2281 ATGCATGGCTTAATGACATCCAAAAACAGAGGTAATGTTTAAATCCATATAGAGTAGAAG 2340
Db 2281 ATGCATGGCTTAATGACATCCAAAAACAGAGGTAATGTTTAAATCCATATAGAGTAGAAG 2340
QY 2341 AAGCCCTGCTATTTAAAAAGACTTTCTTGAGAACTCATAAACTTCTCTGAGAGCCACTGGAAA 2400
Db 2341 AAGCCCTGCTATTTAAAAAGACTTTCTTGAGAACTCATAAACTTCTCTGAGAGCCACTGGAAA 2400
QY 2401 AGCCAAATTAATGTTTAAACAGAGAGTCTTTTAAATCGATCTCATTTATTTTAGAGCTTTTATATA 2460
Db 2401 AGCCAAATTAATGTTTAAACAGAGAGTCTTTTAAATCGATCTCATTTATTTTAGAGCTTTTATATA 2460

Db	241	TGCGAGNTAACGGGAGGGTATCAAGGCCTGTGATGCACCCGTGAATGGCAAATGAAGTA	300
Qy	301	ACAACCTCAAAAAATAAATAGTCATGAAGATCTTGAAAATTGACAACTTACGGTTTTCCG	360
Db	301	ACAACCTCAAAAAATAAATAGTCATGAAGATCTTGAAAATTGACAACTTACGGTTTTCCG	360
Qy	361	GAGAAGCCTTTGGGCTCAATTTGTTGTATAGCTGAGGTTTTAAATTACAACAAGAACGGCTG	420
Db	361	GAGNAGCCTTTGGGCTCAATTTGTTGTATAGCTGAGGTTTTAAATTACAACAAGAACGGCTG	420
Qy	421	CTGATAAATTTAGCACCCGATGTGTTTTAGTAGGCCAGTGCCACATACTTTCTCAGAAAC	480
Db	421	CTGATAAATTTAGCACCCGATGTGTTTTAGTAGGCCAGTGCCACATACTTTCTCAGAAAC	480
Qy	481	CTTTCACATCTTGGTCAAGGTACAACCTGAATCTGTTTTAAGATTTATTAAAGAACTTACCCTG	540
Db	481	CTTTCACATCTTGGTCAAGGTACAACCTGAATCTGTTTTAAGATTTATTAAAGAACTTACCCTG	540
Qy	541	TAAGAAAGCAGTTTTTACTCAACTGCAAAAAAATGTAAAGATGAATATAAAAAAGATCCAAG	600
Db	541	TAAGAAAGCAGTTTTTACTCAACTGCAAAAAAATGTAAAGATGAATATAAAAAAGATCCAAG	600
Qy	601	ATCTCTCATGAGCTTTGGTATCTTAAACCTGACTTAAGGATTTGCTTTGTACATAACA	660
Db	601	ATCTCTCATGAGCTTTGGTATCTTAAACCTGACTTAAGGATTTGCTTTGTACATAACA	660
Qy	661	AGGCAGTATTTGGCAGAAAAGCAGAGTATCAGATCACAAGATGGCTCTCATCTCAGTTTC	720
Db	661	AGGCAGTATTTGGCAGAAAAGCAGAGTATCAGATCACAAGATGGCTCTCATCTCAGTTTC	720
Qy	721	TGGGGACTGCTGTTATGAACAATAATGGAATCTTTTCAGTACCACTCTGGAAGAATCTCAGA	780
Db	721	TGGGGACTGCTGTTATGAACAATAATGGAATCTTTTCAGTACCACTCTGGAAGAATCTCAGA	780
Qy	781	TTTTATCTAGTGAGATTTCTTCCAAAGTGTGATGCAGACCACTCTTTTACCTAGTCTTTCAA	840
Db	781	TTTTATCTAGTGAGATTTCTTCCAAAGTGTGATGCAGACCACTCTTTTACCTAGTCTTTCAA	840
Qy	841	CACAGAAAGAAGTTTTCATCTTTCATAAACAGTGCAGCAGTACATCAAAAAAGATATCTTAA	900
Db	841	CACAGAAAGAAGTTTTCATCTTTCATAAACAGTGCAGCAGTACATCAAAAAAGATATCTTAA	900
Qy	901	AGTTAAATCCGACATCATTAACAATCTGAATGSCCTTAAAGGAATCTACTGTTTGTATCCTG	960
Db	901	AGTTAAATCCGACATCATTAACAATCTGAATGSCCTTAAAGGAATCTACTGTTTGTATCCTG	960
Qy	961	TTTTCTTTCTGAAAATCGATGTTCCCTACAGCTGATGTGTATTAATTCCTGATAATTAACAC	1020
Db	961	TTTTCTTTCTGAAAATCGATGTTCCCTACAGCTGATGTGTATTAATTCCTGATAATTAACAC	1020
Qy	1021	AAAGCCAGTATTAATTAACAAATTAAGGAATCTGTTTTAATTCGCTTCTGAAAAATCTGATGA	1080
Db	1021	AAAGCCAGTATTAATTAACAAATTAAGGAATCTGTTTTAATTCGCTTCTGAAAAATCTGATGA	1080
Qy	1081	CGACTTGTATGGACCAATTAACCTAGTACAAATCTTATGAAAAATTAATAACAGATGTTT	1140
Db	1081	CGACTTGTATGGACCAATTAACCTAGTACAAATCTTATGAAAAATTAATAACAGATGTTT	1140
Qy	1141	CCGAGCTGCAATCGTTCTTTAGTAAACACAGCAGAAAAGATGCTGTTTTTAATAAAGTGG	1200
Db	1141	CCGAGCTGCAATCGTTCTTTAGTAAACACAGCAGAAAAGATGCTGTTTTTAATAAAGTGG	1200
Qy	1201	AATCATCTGGAAAGAAATTTATTCAAATGTTGATCTTCAGTCAATTCCTATCCAAATGATA	1260
Db	1201	AATCATCTGGAAAGAAATTTATTCAAATGTTGATCTTCAGTCAATTCCTATCCAAATGATA	1260
Qy	1261	TGCATAATGATCAATCTCGAAAAAAACACTGATGATGTTGTTTAAATCAACAGATAAGTATG	1320
Db	1261	TGCATAATGATCAATCTCGAAAAAAACACTGATGATGTTGTTTAAATCAACAGATAAGTATG	1320
Qy	1321	GTGACTTTGGTTATGGTCAATTTGTAGTAGTGAATTTTCTAAACATTTGATTAACACACTAAGA	1380

1321	Db	GTGACCTTTGGTTAATGCTCATTTGTAGTAGTGAAATTTCTTAACATTGATAAAAAACACTAAGA	1338
1381	Qy	ATGCATTTTCAGGACATTTTCAATCAGTAATAATGTATCATGGGAGAACTCTCAGACGGAAATATA	1440
1381	Db	ATGCATTTTCAGGACATTTTCAATCAGTAATAATGTATCATGGGAGAACTCTCAGACGGAAATATA	1440
1441	Qy	GTAAACCTTGCTTTTATAAGTTCCGTTTAAGCACACCCAGTCAGAAAAATGGCAATAAAGACC	1500
1441	Db	GTAAACCTTGCTTTTATAAGTTCCGTTTAAGCACACCCAGTCAGAAAAATGGCAATAAAGACC	1500
1501	Qy	ATATAGATCAGAGTGGGAAAAATGAGGAAGACGAGCTTTTGAAAACTCTTCGGAATTTT	1560
1501	Db	ATATAGATCAGAGTGGGAAAAATGAGGAAGACGAGCTTTTGAAAACTCTTCGGAATTTT	1560
1561	Qy	CTGCAGATGATGGACAGCGGGAAATATATCTTTAAAAATTCAGTGGGAGAGAAATATTCGAAC	1620
1561	Db	CTGCAGATGATGGACAGCGGGAAATATATCTTTAAAAATTCAGTGGGAGAGAAATATTCGAAC	1620
1621	Qy	CTGTGAAAAATTTTAGTGCCTGAAAAAAGTTTACCATGTAAGTAAGTAATAATAATTTATC	1680
1621	Db	CTGTGAAAAATTTTAGTGCCTGAAAAAAGTTTACCATGTAAGTAAGTAATAATAATTTATC	1680
1681	Qy	CAATCCCTGAAACAAATGAAATCTTAAATGAAGATTCATGTAACAAAAATCAAAATGTAATAG	1740
1681	Db	CAATCCCTGAAACAAATGAAATCTTAAATGAAGATTCATGTAACAAAAATCAAAATGTAATAG	1740
1741	Qy	ATAATAAATCTGAAAAAGTTTACAGCTTATGATTTTACTTTAGCAATCGAGTAAATCAAGAAC	1800
1741	Db	ATAATAAATCTGAAAAAGTTTACAGCTTATGATTTTACTTTAGCAATCGAGTAAATCAAGAAC	1800
1801	Qy	CCATGTCAGCAAGTGCCTCTTTTGTGTTCAAGATCATCGTCTCAGTTTCTCATAGAAAATC	1860
1801	Db	CCATGTCAGCAAGTGCCTCTTTTGTGTTCAAGATCATCGTCTCAGTTTCTCATAGAAAATC	1860
1861	Qy	CTAAGACTAGTTTAGAGGATGCAACACTAATAATTGAAGAACTGTGGAAGACATTGTGATG	1920
1861	Db	CTAAGACTAGTTTAGAGGATGCAACACTAATAATTGAAGAACTGTGGAAGACATTGTGATG	1920
1921	Qy	AAGAGGAAAAACTCAAAATATCGAGAGAGGCTACTTAAAGACTTGGACAGGATACAAATAGTC	1980
1921	Db	AAGAGGAAAAACTCAAAATATCGAGAGAGGCTACTTAAAGACTTGGACAGGATACAAATAGTC	1980
1981	Qy	AAATGAAGAGAGCCATTTGAACAGAGGTCACAAAATGTCTATAAAGATGGCAGAAAAAAGA	2040
1981	Db	AAATGAAGAGAGCCATTTGAACAGAGGTCACAAAATGTCTATAAAGATGGCAGAAAAAAGA	2040
2041	Qy	TAAACCCACAGCGCATGGAAATTTGGCCGAGAGCAAGTTTAAAAACCTCATTTACTA	2100
2041	Db	TAAACCCACAGCGCATGGAAATTTGGCCGAGAGCAAGTTTAAAAACCTCATTTACTA	2100
2101	Qy	ATCAACCAAACTTTGATGAACTCCCTCAGTCCCAAAATTTGAAAAAGAGGAGTCAAAATA	2160
2101	Db	ATCAACCAAACTTTGATGAACTCCCTCAGTCCCAAAATTTGAAAAAGAGGAGTCAAAATA	2160
2161	Qy	TTAAAAATGGTACAGATCCCCCTTTTCTATGAAAAAATTTAAAAATTTTAAAGAACAAA	2220
2161	Db	TTAAAAATGGTACAGATCCCCCTTTTCTATGAAAAAATTTAAAAATTTTAAAGAACAAA	2220
2221	Qy	ACAAAGTTGACTTGAAGAGAGGATGAACTTTGCTGTGATCCCAATCTCAGGTTTCCTG	2280
2221	Db	ACAAAGTTGACTTGAAGAGAGGATGAACTTTGCTGTGATCCCAATCTCAGGTTTCCTG	2280
2281	Qy	ATGCATGGCTAATGCATCCAAAAACAGAGGTAAATGTTATTAATCCATATAGGTAGAG	2340
2281	Db	ATGCATGGCTAATGCATCCAAAAACAGAGGTAAATGTTATTAATCCATATAGGTAGAG	2340
2341	Qy	AAGCCCTCTATTTTAAAAAGACTTCTTGAGAAATCATAAATTTTCTCCTCAGAGCCACTGGAAA	2400
2341	Db	AAGCCCTCTATTTTAAAAAGACTTCTTGAGAAATCATAAATTTTCTCCTCAGAGCCACTGGAAA	2400
2401	Qy	AGCCAAATTTATGTTAAACAGAGAGTCTTTTTTAATGGATCTCATTTATTAGACGTTTTATATA	2460
2401	Db	AGCCAAATTTATGTTAAACAGAGAGTCTTTTTTAATGGATCTCATTTATTAGACGTTTTATATA	2460

GenCore version 5.1.6
Copyright (C) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model
Run on: December 26, 2005, 11:37:28 ; Search time 244 Seconds
(without alignments)
6514.585 Million cell updates/sec

Title: US-10-079-429a-3
Perfect score: 3063
Sequence: 1 ggcacgagtgctgctgcg.....aacgtaataaactaataac 3063

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4168288 seqs, 259477437 residues

Total number of hits satisfying chosen parameters: 8336576

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications NA New:
1: /cgn2_6/prodata/1/pubpna/US08_NEW_PUB.seq:
2: /cgn2_6/prodata/1/pubpna/US06_NEW_PUB.seq:
3: /cgn2_6/prodata/1/pubpna/US07_NEW_PUB.seq:
4: /cgn2_6/prodata/1/pubpna/US09_NEW_PUB.seq:
5: /cgn2_6/prodata/1/pubpna/US10_NEW_PUB.seq:
6: /cgn2_6/prodata/1/pubpna/US11_NEW_PUB.seq:
7: /cgn2_6/prodata/1/pubpna/US12_NEW_PUB.seq:
8: /cgn2_6/prodata/1/pubpna/US13_NEW_PUB.seq:
9: /cgn2_6/prodata/1/pubpna/US14_NEW_PUB.seq:
10: /cgn2_6/prodata/1/pubpna/US15_NEW_PUB.seq:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3063	100.0	3063	7	US-11-188-743-8 Sequence 8, Appli
2	99.8	3.3	2589	7	US-11-128-420-3 Sequence 3, Appli
3	99.8	3.3	2771	7	US-11-188-743-7 Sequence 7, Appli
4	94.2	3.1	1408	7	US-11-188-743-12 Sequence 12, Appli
5	91.4	3.0	3056	7	US-11-188-743-6 Sequence 6, Appli
6	84.8	2.8	402	7	US-11-128-420-5 Sequence 5, Appli
7	84	2.7	426	7	US-11-188-743-11 Sequence 11, Appli
8	80.4	2.6	1785	7	US-11-188-743-13 Sequence 13, Appli
9	76.8	2.5	2484	7	US-11-188-743-10 Sequence 10, Appli
10	75.2	2.5	3218	7	US-11-188-743-5 Sequence 5, Appli
11	72.4	2.4	2340	7	US-11-128-420-4 Sequence 4, Appli
12	55	1.8	3000	7	US-11-194-246-150 Sequence 150, App
13	53.4	1.7	441	7	US-11-128-420-6 Sequence 6, Appli
14	53	1.7	1974	6	US-10-467-657-4781 Sequence 4781, Ap
15	52	1.7	795	7	US-11-188-743-14 Sequence 14, Appli
16	49.4	1.6	1284	7	US-11-074-176-343 Sequence 343, App
17	49.4	1.6	1289	7	US-11-074-176-161 Sequence 161, App
18	48.6	1.6	134499	7	US-11-117-187-192 Sequence 192, App
19	45.8	1.5	305312	6	US-10-995-561-13236 Sequence 13236, A
20	43	1.4	190276	6	US-10-661-966-1 Sequence 1, Appli
21	42.6	1.4	2088	6	US-10-793-626-3463 Sequence 3463, Ap
22	42.4	1.4	191797	7	US-11-121-086-13 Sequence 13, Appli
23	41.6	1.4	600	6	US-10-750-185-20965 Sequence 20965, A

ALIGNMENTS

RESULT 1

US-11-188-743-8
; Sequence 8, Application US/11188743
; Publication NO. US20050272140A1
; GENERAL INFORMATION:
; APPLICANT: Nicolaides, Nicholas
; APPLICANT: Sassi, Philip
; APPLICANT: Kinzier, Kenneth
; APPLICANT: Grasso, Luigi
; APPLICANT: Vogelstein, Bert
; TITLE OF INVENTION: Methods for generating hypermutable
; FILE REFERENCE: 01107.00097
; CURRENT APPLICATION NUMBER: US/11/188,743
; CURRENT FILING DATE: 2005-07-26
; PRIOR APPLICATION NUMBER: US/10/641,068
; PRIOR FILING DATE: 2003-08-15
; PRIOR APPLICATION NUMBER: US/09/788,657
; PRIOR FILING DATE: 2001-02-21
; PRIOR APPLICATION NUMBER: 60/184,336
; PRIOR FILING DATE: 2000-02-23
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 8
; LENGTH: 3063
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-188-743-8

Query Match 100.0%; Score 3063; DB 7; Length 3063;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 3063; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GGCACGAGTGGCTGCTGGCTAGTGGATGGTAATTCCTCGCTAGCAGCAAG 60
DB 1 GGCACGAGTGGCTGCTGGCTAGTGGATGGTAATTCCTCGCTAGCAGCAAG 60
QY 61 CTGCTCTGTAAAGCGAAATGAACAATTCCTCGCTAGCAGTTCACCTCTTCAA 120
DB 61 CTGCTCTGTAAAGCGAAATGAACAATTCCTCGCTAGCAGTTCACCTCTTCAA 120
QY 121 GTTCTCAGATCATCTCTCGGTGGTGTGTAAGAGCTTATTGAAAACCTCTTGG 180
DB 121 GTTCTCAGATCATCTCTCGGTGGTGTGTAAGAGCTTATTGAAAACCTCTTGG 180
QY 181 ATGCTGGTCCACAGCGTAGATGTTAAACTGGAGACTATGGATTGTAAATTTGAGG 240

181 ATGCTGGTCCCAAGCGTAGATGTTAAACTGGAGAACTATGATTTGATATAAAATGAGG 240
241 TCGGAGATAACGGGAGGGGTATCAAGGCTGTGATGCACCTGTAATGGCAATGAAGTACT 300
241 TCGGAGATAACGGGAGGGGTATCAAGGCTGTGATGCACCTGTAATGGCAATGAAGTACT 300
301 ACACCTCAAAAATAAATAGTCAATGAAGATCTTGAAGAAATTTGACAACTTACCGGTTTCGGT 360
301 ACACCTCAAAAATAAATAGTCAATGAAGATCTTGAAGAAATTTGACAACTTACCGGTTTCGGT 360
361 GAGAAGCCTTGGGCTCAATTTGTTGTATAGCTGAGGTTTAAATTAACAACAAGACGGCTG 420
361 GAGAAGCCTTGGGCTCAATTTGTTGTATAGCTGAGGTTTAAATTAACAACAAGACGGCTG 420
421 CTGATAAATTTAGCACCCAGTATGTTTATAGATGGCAGTGGCCACATATCTTTCTCAGAAC 480
421 CTGATAAATTTAGCACCCAGTATGTTTATAGATGGCAGTGGCCACATATCTTTCTCAGAAC 480
481 CTTCCATCTTGGTCAAGGTACAACTGTAACCTGCTTAAAGATTAATTTAAGAAATCTACCTG 540
481 CTTCCATCTTGGTCAAGGTACAACTGTAACCTGCTTAAAGATTAATTTAAGAAATCTACCTG 540
541 TAAGAAAGCAGTTTACTCAACTGCAAAAATAATGTAAGATGAATTAATAAAGATCCAAAG 600
541 TAAGAAAGCAGTTTACTCAACTGCAAAAATAATGTAAGATGAATTAATAAAGATCCAAAG 600
601 ATCTCCTCATGAGCTTTGGTATCTTTAAACCTGACTTAAGGATGTCTTTGTACATAACA 660
601 ATCTCCTCATGAGCTTTGGTATCTTTAAACCTGACTTAAGGATGTCTTTGTACATAACA 660
661 AGGCAATTTTGGCAAGAGAGTATCAGATCACAAGATGCTCTCATGTCAAGTTC 720
661 AGGCAATTTTGGCAAGAGAGTATCAGATCACAAGATGCTCTCATGTCAAGTTC 720
721 TGGGACCTGCTGTATGAACAATATGGAATCTCTTCAAGTACCACTCTGAAGAATCTCAGA 780
721 TGGGACCTGCTGTATGAACAATATGGAATCTCTTCAAGTACCACTCTGAAGAATCTCAGA 780
781 TTTTATCTCAGTGGATTTCTTCAAAAGTGTGATGCAGACCACTCTTTCACTAGTCTTTCAA 840
781 TTTTATCTCAGTGGATTTCTTCAAAAGTGTGATGCAGACCACTCTTTCACTAGTCTTTCAA 840
841 CACCGAAGAGATTTTCATCTTCAAAAAGTGTGATGCAGACCACTCTTTCACTAGTCTTTCAA 900
841 CACCGAAGAGATTTTCATCTTCAAAAAGTGTGATGCAGACCACTCTTTCACTAGTCTTTCAA 900
901 AGTTAATCCGACATCAATTAACAATCTGAAATGCTTAAAGGAATCTACTCGTTTGTATCCTG 960
901 AGTTAATCCGACATCAATTAACAATCTGAAATGCTTAAAGGAATCTACTCGTTTGTATCCTG 960
961 TTTTCTTTCTGAAAATCGATGTTCTTACAGCTGATGTGATGTAATTTAAACACAGATA 1020
961 TTTTCTTTCTGAAAATCGATGTTCTTACAGCTGATGTGATGTAATTTAAACACAGATA 1020
1021 AAAGCCAAAGTATTTAACAATAAAGGAATCTGTTTAAATGCTCTTGAAGAAATCTGATGA 1080
1021 AAAGCCAAAGTATTTAACAATAAAGGAATCTGTTTAAATGCTCTTGAAGAAATCTGATGA 1080
1081 CGACTTGTATGGACCATTTACCTAGTACAAATTTCTTATGAAGAAATTAATAACAGATGTTT 1140
1081 CGACTTGTATGGACCATTTACCTAGTACAAATTTCTTATGAAGAAATTAATAACAGATGTTT 1140
1141 CCGCAGCTGACATCGTTCTTAGTAAACAGCAGAAACAGATGTGCTTTTAAATAAGTGG 1200
1141 CCGCAGCTGACATCGTTCTTAGTAAACAGCAGAAACAGATGTGCTTTTAAATAAGTGG 1200
1201 AATCATCTGGAAGAAATTTAATAATTTGATACCTTCAAGTCTTCAATCCAAATGATA 1260
1201 AATCATCTGGAAGAAATTTAATAATTTGATACCTTCAAGTCTTCAATCCAAATGATA 1260
1261 TGCATAATGATGATCTGGAAAAACACATGATGTTTAAATCACCAGATAAGTATTTG 1320

1261 TGCATAATGATGATCTGGAAAAACACATGATGATGTTTAAATCACCAGATAAGTATTTG 1320
1321 GTGACTTTGGTATGCTCATGTTAGTAAATTTCTTAAACATGATATAAAACACTAAAGA 1380
1321 GTGACTTTGGTATGCTCATGTTAGTAAATTTCTTAAACATGATATAAAACACTAAAGA 1380
1381 ATGCATTTTCCAGGACATTTCAATGATGATATCATGGGAGAACTCTCTCAGACGGAATATA 1440
1381 ATGCATTTTCCAGGACATTTCAATGATGATATCATGGGAGAACTCTCTCAGACGGAATATA 1440
1441 GTAAAACTTGTGTTTATAAGTTCCGTTAAGCACACCCAGTCAAGAAATGCGAATAAAGACC 1500
1441 GTAAAACTTGTGTTTATAAGTTCCGTTAAGCACACCCAGTCAAGAAATGCGAATAAAGACC 1500
1501 ATATAGATGAGATGGGGGAAAATGAGGAAAGACAGGCTCTTGAAAACTCTTTCGGAAAAATTT 1560
1501 ATATAGATGAGATGGGGGAAAATGAGGAAAGACAGGCTCTTGAAAACTCTTTCGGAAAAATTT 1560
1561 CTGCAGATGAGTGGGAGCAGGGGAAAATATACTTAAAAATTCAGTGGGAGAGAAATATTGAAC 1620
1561 CTGCAGATGAGTGGGAGCAGGGGAAAATATACTTAAAAATTCAGTGGGAGAGAAATATTGAAC 1620
1621 CTGTGAAAAATTTTGTAGTCTGAAAAAAGTTTACCATGTAAAGTAAATTAATAATTATC 1680
1621 CTGTGAAAAATTTTGTAGTCTGAAAAAAGTTTACCATGTAAAGTAAATTAATAATTATC 1680
1681 CAATCCCTGAAACAAATGAATCTTAAATGAAGATTCATGTAAACAAAAATCAAAATGTAATAG 1740
1681 CAATCCCTGAAACAAATGAATCTTAAATGAAGATTCATGTAAACAAAAATCAAAATGTAATAG 1740
1741 ATAATAAATCTGGAAAAAGTTACAGCTTATGATTTACTTAGCAATCGAGTAAATCAAGAAAC 1800
1741 ATAATAAATCTGGAAAAAGTTACAGCTTATGATTTACTTAGCAATCGAGTAAATCAAGAAAC 1800
1801 CCAATGTACAGAAAGTCTCTTTTGTTCAGATCATCGCTCCTCAGTTTCTCATAGAAAAATC 1860
1801 CCAATGTACAGAAAGTCTCTTTTGTTCAGATCATCGCTCCTCAGTTTCTCATAGAAAAATC 1860
1861 CTAAGCATGTTTACAGGATGCAACACTACAAATTTGAAGAACTGTGGAGACATTTGAGTG 1920
1861 CTAAGCATGTTTACAGGATGCAACACTACAAATTTGAAGAACTGTGGAGACATTTGAGTG 1920
1921 AAGAGGAAAACTGAAAATATGAAGAGAGGCTTCTTAAAGACTTTGGAAACGATACAAATAGTC 1980
1921 AAGAGGAAAACTGAAAATATGAAGAGAGGCTTCTTAAAGACTTTGGAAACGATACAAATAGTC 1980
1981 AAATGAAGAGAGCCATTGAACAGGAGTCAAAATGTCACTAAAGATGGCAGAAAAAAGA 2040
1981 AAATGAAGAGAGCCATTGAACAGGAGTCAAAATGTCACTAAAGATGGCAGAAAAAAGA 2040
2041 TAAAAACCCACAGCGCATGGAAATTTGGCCAGAGCAAAAGTAAAAAACTCATTTATCTA 2100
2041 TAAAAACCCACAGCGCATGGAAATTTGGCCAGAGCAAAAGTAAAAAACTCATTTATCTA 2100
2101 ATCAACCCAAAACTTGATGAATCTCTTCAAGTCCCAAATTTGAAAAAAGAGAGGTCAAAAA 2160
2101 ATCAACCCAAAACTTGATGAATCTCTTCAAGTCCCAAATTTGAAAAAAGAGAGGTCAAAAA 2160
2161 TTAATAATGTTACAGATCCCTCTTTCTATGAAAACTTAAAAATTAATTTTAAAGAAACAAA 2220
2161 TTAATAATGTTACAGATCCCTCTTTCTATGAAAACTTAAAAATTAATTTTAAAGAAACAAA 2220
2221 ACAAAAGTTGACTTTAGAGAGAGGATGAACCTTGTGATCCCAATCTCAGGTTTCTTG 2280
2221 ACAAAAGTTGACTTTAGAGAGAGGATGAACCTTGTGATCCCAATCTCAGGTTTCTTG 2280
2281 ATGCATGCTGTAATGATCCAAAAACAGAGGTAATGTTTAAATTCATATAGAGTGAAG 2340
2281 ATGCATGCTGTAATGATCCAAAAACAGAGGTAATGTTTAAATTCATATAGAGTGAAG 2340
2341 AAGCCCTGCTATTTAAAGACTCTTGTGAGAAATCATAACTTCTTCAGAGGCCACTCGAAA 2400
2341 AAGCCCTGCTATTTAAAGACTCTTGTGAGAAATCATAACTTCTTCAGAGGCCACTCGAAA 2400

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 26, 2005, 14:40:21 ; Search time 10058 Seconds
(without alignments)
17310.747 Million cell updates/sec

Title: US-10-079-429A-3
Perfect score: 3063
Sequence: 1 ggcagagtggtgtgtgcg.....aacgtaataaactaataac 3063

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 5883141 seqs, 28421725653 residues
Word size : 0
Total number of hits satisfying chosen parameters: 11766282

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : GenEmbl.*

1: gb_ba.*
2: gb_in.*
3: gb_env.*
4: gb_om.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pr.*
9: gb_ro.*
10: gb_sta.*
11: gb_sy.*
12: gb_un.*
13: gb_vi.*
14: gb_htg.*
15: gb_pl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3063	100.0	3063	6	BD181098 Human DNA
2	3063	100.0	3063	6	CS020326 Sequence
3	3063	100.0	3063	6	CS056082 Sequence
4	3063	100.0	3063	6	AR217665 Sequence
5	3063	100.0	3063	6	AR342820 Sequence
6	3063	100.0	3063	6	AR380890 Sequence
7	3063	100.0	3063	6	AR382023 Sequence
8	3063	100.0	3063	6	AR437149 Sequence
9	3063	100.0	3063	6	AR540769 Sequence
10	3063	100.0	3063	6	AR592873 Sequence
11	3063	100.0	3063	6	AX214170 Sequence
12	3063	100.0	3063	6	AX234586 Sequence
13	3063	100.0	3063	6	AX775137 Sequence
14	3063	100.0	3063	8	UI3695 Human homol
15	3004	98.1	3239	6	CS130851 Sequence
16	2830	92.4	2830	8	BC096332 Homo sapi
17	2732	89.2	2830	8	BC096330 Homo sapi
18	2102	68.6	2882	8	AB102875 Homo sapi

19	1979	64.6	2856	6	CQ720889 Sequence
20	1855	60.9	2343	8	BC096331 Homo sapi
21	1856	60.6	2313	8	AB102870 Homo sapi
22	1624	53.0	2252	8	AY540750 Homo sapi
23	1392	45.4	2004	8	AB102869 Homo sapi
24	1391	45.4	2817	8	AY540751 Homo sapi
25	1159	37.8	1668	8	AB102872 Homo sapi
26	892	29.1	104140	8	AC008122 Homo sapi
27	892	29.1	106234	8	AY267352 Homo sapi
28	790	25.8	155122	14	AC141847
29	790	25.8	212131	14	AC142554
30	701	22.9	705	8	AB102876 Homo sapi
31	583	19.0	747	8	AB102874 Homo sapi
32	592	19.0	588	8	AB102877 Homo sapi
33	582	19.0	591	8	AB102873 Homo sapi
34	493	16.1	1906	8	BC008410 Homo sapi
35	486	15.9	731	10	G67580
36	442	14.4	965	8	BC084548 Homo sapi
37	421	13.7	501	11	BT006947 Homo sapi
38	421	13.7	501	11	AY335564 Synthetic
39	421	13.7	501	11	BT007647 Synthetic
40	233	7.6	419	6	CQ680441 Sequence
41	211	6.9	188715	14	AC141850
42	187	6.1	327	6	AR255689 Sequence
43	187	6.1	327	6	AR399511 Sequence
44	167	5.5	389	10	G67582 csnpmsl-pc
45	154	5.0	559	6	CQ516182 Sequence

ALIGNMENTS

RESULT 1	BD181098	Human DNA mismatch repair proteins.	3063 bp	DNA	linear	PAT 15-MAY-2003
LOCUS	BD181098					
DEFINITION	BD181098					
ACCESSION	BD181098.1	GI:30792016				
VERSION	BD181098.1	GI:30792016				
KEYWORDS	JP 2002325588-A/2.					
SOURCE	Homo sapiens (human)					
ORGANISM	Homo sapiens					
REFERENCE	1 (bases 1 to 3063)					
AUTHORS	Haseltine,W.A., Ruben,S.M., Wei,Y.F., Adams,M.D., Fleischmann,R.D., Fraser,C.M., Fuldner,R.A., Kirkness,E.F. and Rosen,C.A.					
TITLE	Human DNA mismatch repair proteins					
JOURNAL	Patent: JP 2002325588-A 2 12-NOV-2002;					
COMMENT	HUMAN GENOME SCIENCES INC					
	OS Homo sapiens (human)					
	PN JP 2002325588-A/2					
	PD 12-NOV-2002					
	PF 25-JAN-2002 JP 2002016830					
	PR 27-JAN-1994 US 08/187757;16-MAR-1994 US 08/210143 PR					
	23-AUG-1994 US 08/294312					
	PI WILLIAM A HASELTINE,STEVEN M RUBEN,YING FEI WEI,MARK D ADAMS,					
	PI ROBERT D FLEISCHMANN,CLAIRE M FRASER,REBECCA A FULDNER,EWEN F					
	PI KIRKNESS,					
	PI CRAIG A ROSEN					
	PC C12N15/09,C07K14/47,C12P21/02,C12Q1/68//((C12P21/02,C12R1:19),					
	PC C12N15/00					
	CC Human DNA mismatch repair proteins					
	FH Key Location/Qualifiers					
	FT CDS					
FEATURES	1..3063					
source	/organism="Homo sapiens"					
	/mol_type="genomic DNA"					
	/db_xref="taxon:9606"					
ORIGIN						
Query Match	100.0%;	Score	3063;	DB	6;	Length 3063;

Best Local Similarity 100.0%; Pred. No. 0; Matches 3063; Conservative 0; Mismatches 0; Indels 0; Gaps 0;			
Qy	1	GGCAGAGTGGCTGCTTGGCGCTAGTGGATGTAATGGCTCGCTCGCGTAGCAGCAAG	60
Db	1	GGCAGAGTGGCTGCTTGGCGCTAGTGGATGTAATGGCTCGCTCGCGTAGCAGCAAG	60
Qy	61	CTGCTCTGTAAAGCGAAATGAAACAATGGCTCGCGCAACAGTTCGACTCCTTTCAA	120
Db	61	CTGCTCTGTAAAGCGAAATGAAACAATGGCTCGCGCAACAGTTCGACTCCTTTCAA	120
Qy	121	GTTCTCAGATCATCTTCGGTGGTGGTGTAAAGAGCTTATGAAAACTCCTTGG	180
Db	121	GTTCTCAGATCATCTTCGGTGGTGGTGTAAAGAGCTTATGAAAACTCCTTGG	180
Qy	181	ATGCTGGTCCCAAGCGTAGATGTTAAACTGGAGAACTATGGATTTGATAAAATGGAG	240
Db	181	ATGCTGGTCCCAAGCGTAGATGTTAAACTGGAGAACTATGGATTTGATAAAATGGAG	240
Qy	241	TGCGAGATAACGGGGAGGTATCAAGGCTGTTGATGCACCTGTATGGCAATGAAGTACT	300
Db	241	TGCGAGATAACGGGGAGGTATCAAGGCTGTTGATGCACCTGTATGGCAATGAAGTACT	300
Qy	301	ACACCTCAAAAATAAATAGTCAATGTAAGATCTTGAAAAATTTGACACTTACGGTTTTCGTG	360
Db	301	ACACCTCAAAAATAAATAGTCAATGTAAGATCTTGAAAAATTTGACACTTACGGTTTTCGTG	360
Qy	361	GAGAGCCTTGGGTCAAATTTGTTGATAGCTGAGGTTTTTAATTAACAACAAGACGGCTG	420
Db	361	GAGAGCCTTGGGTCAAATTTGTTGATAGCTGAGGTTTTTAATTAACAACAAGACGGCTG	420
Qy	421	CTGATAAATTTAGCAACCCAGTATGTTTTAGATGGCAGTGGCCACATACCTTCTCAGAAAC	480
Db	421	CTGATAAATTTAGCAACCCAGTATGTTTTAGATGGCAGTGGCCACATACCTTCTCAGAAAC	480
Qy	481	CTTCACATCTGGTCAAGGTCAACTGTAAGTCTTTTAAGATTTATTAAGATTTACCTT	540
Db	481	CTTCACATCTGGTCAAGGTCAACTGTAAGTCTTTTAAGATTTATTAAGATTTACCTT	540
Qy	541	TAAGAAAGCAGTTTTTACTCAACTGCAAAAAATGTAAGATGAAATTAAGAAATCCAAAG	600
Db	541	TAAGAAAGCAGTTTTTACTCAACTGCAAAAAATGTAAGATGAAATTAAGAAATCCAAAG	600
Qy	601	ATCTCTCATGAGCTTTGGTATCCTTTAAACCTGACTTAAGGATGTCTTTGTACATAACA	660
Db	601	ATCTCTCATGAGCTTTGGTATCCTTTAAACCTGACTTAAGGATGTCTTTGTACATAACA	660
Qy	661	AGGCAGTTATTTGGCAGAAAGCAGAGTATCAGATCACAAGATGGCTCTCATGTCAGTTTC	720
Db	661	AGGCAGTTATTTGGCAGAAAGCAGAGTATCAGATCACAAGATGGCTCTCATGTCAGTTTC	720
Qy	721	TGSGGACTGCTGTATGAAACAATATGAAATCCTTTTCAGTACCACTCTGAAGAACTCAGA	780
Db	721	TGSGGACTGCTGTATGAAACAATATGAAATCCTTTTCAGTACCACTCTGAAGAACTCAGA	780
Qy	781	TTTATCTCAGTGGATTTCTTCCAAAGTGTGATGCAGACCACTCTTTCACATAGTCTTTCAA	840
Db	781	TTTATCTCAGTGGATTTCTTCCAAAGTGTGATGCAGACCACTCTTTCACATAGTCTTTCAA	840
Qy	841	CACGAGAAAGAGTTTCATCTTCAATAACAGTGCAGCAGTACATCAAAAAGATATCTTAA	900
Db	841	CACGAGAAAGAGTTTCATCTTCAATAACAGTGCAGCAGTACATCAAAAAGATATCTTAA	900
Qy	901	AGTTAATCCACATCATTTACAACTGAAATGCGCTTAAAGGAATCTACTCGTTTGTATCCTG	960
Db	901	AGTTAATCCACATCATTTACAACTGAAATGCGCTTAAAGGAATCTACTCGTTTGTATCCTG	960
Qy	961	TTTTCTTCTGAAAAATCGATGTTCTTCAAGCTGATGTTGATGTAATTTAAACCCAGATA	1020
Db	961	TTTTCTTCTGAAAAATCGATGTTCTTCAAGCTGATGTTGATGTAATTTAAACCCAGATA	1020
Qy	1021	AAAGCCAAAGTATTATTACAAAATAAGGAATCTGTTTAAATGCTCTTGAAGAACTGATGA	1080
Db	1021	AAAGCCAAAGTATTATTACAAAATAAGGAATCTGTTTAAATGCTCTTGAAGAACTGATGA	1080
Qy	1081	CGACTTGTATGGACCAATACCTAGTACAATCTTATGAAAAATAAATAAACAGATGTTT	1140
Db	1081	CGACTTGTATGGACCAATACCTAGTACAATCTTATGAAAAATAAATAAACAGATGTTT	1140
Qy	1141	CCGAGCTGCAATCGTTCTTTAGTAAAAACAGCAGAAAAACAGATGTGCTTTTAAATGAGTGG	1200
Db	1141	CCGAGCTGCAATCGTTCTTTAGTAAAAACAGCAGAAAAACAGATGTGCTTTTAAATGAGTGG	1200
Qy	1201	AATCATCTGGAAGAATTTATTTCAATGTGATATCTTCAATTCATTCATTCCTTCAAAATGATA	1260
Db	1201	AATCATCTGGAAGAATTTATTTCAATGTGATATCTTCAATTCATTCATTCCTTCAAAATGATA	1260
Qy	1261	TGCATAATGATGAATCTGGAAGAAAAACACTGATGATGTTTAAATCACAGATAAGTATTTG	1320
Db	1261	TGCATAATGATGAATCTGGAAGAAAAACACTGATGATGTTTAAATCACAGATAAGTATTTG	1320
Qy	1321	GTGACTTTGGTTATGGTCAITTTAGTGTGAAAAATTTCTAAATGATTAATAAACACTTAAGA	1380
Db	1321	GTGACTTTGGTTATGGTCAITTTAGTGTGAAAAATTTCTAAATGATTAATAAACACTTAAGA	1380
Qy	1381	ATGCAATTTCAAGACATTTCAATGATGATGTAATCTCATGGAGAACTCTCAGACGGAAATA	1440
Db	1381	ATGCAATTTCAAGACATTTCAATGATGATGTAATCTCATGGAGAACTCTCAGACGGAAATA	1440
Qy	1441	GTAAACCTTTGTTTATTAAGTTCCGTTAAGCACACCCAGTCAGAAAAATGCAATAAAGACC	1500
Db	1441	GTAAACCTTTGTTTATTAAGTTCCGTTAAGCACACCCAGTCAGAAAAATGCAATAAAGACC	1500
Qy	1501	ATATAGATGAGAGTGGGAAAAATGAGGAAGACGAGGCTTTGAAAACTCTTCGGAAAAATTT	1560
Db	1501	ATATAGATGAGAGTGGGAAAAATGAGGAAGACGAGGCTTTGAAAACTCTTCGGAAAAATTT	1560
Qy	1561	CTGCAAGATGAGTGGAGCAGGGGAAATATATCTTAAATAATTCAGTGGGAGAGAAATATTGAAC	1620
Db	1561	CTGCAAGATGAGTGGAGCAGGGGAAATATATCTTAAATAATTCAGTGGGAGAGAAATATTGAAC	1620
Qy	1621	CTGTGAAAAATTTTAGTGCTGCTGAAAAAGTTTACCATGTAAAGTAAAGTAAATTAATTAATC	1680
Db	1621	CTGTGAAAAATTTTAGTGCTGCTGAAAAAGTTTACCATGTAAAGTAAAGTAAATTAATTAATC	1680
Qy	1681	CAATCCCTGAAACAAATGAAATCTTAATGAAGATTCATGTAAACAAAAATCAATGTAATAG	1740
Db	1681	CAATCCCTGAAACAAATGAAATCTTAATGAAGATTCATGTAAACAAAAATCAATGTAATAG	1740
Qy	1741	ATAATAAATCTGAAAAAGTTACAGCTTATGATTTACTTTAGCAATCGAGTAATCAAGAAAC	1800
Db	1741	ATAATAAATCTGAAAAAGTTACAGCTTATGATTTACTTTAGCAATCGAGTAATCAAGAAAC	1800
Qy	1801	CCATGTCAGCAAGTGTCTTTTGTGTTTCAAGATCATGCTCTCTCATAGAAAAATC	1860
Db	1801	CCATGTCAGCAAGTGTCTTTTGTGTTTCAAGATCATGCTCTCTCATAGAAAAATC	1860
Qy	1861	CTAAGACTAGTTTAGAGGATGCAACATCAAAATTTGAAGAACTGTGGAGACATTCAGTG	1920
Db	1861	CTAAGACTAGTTTAGAGGATGCAACATCAAAATTTGAAGAACTGTGGAGACATTCAGTG	1920
Qy	1921	AAGAGAAAAAATCTGAAAAATGAGAGAGGCTACTAAAGACTTTGGAACCATCAATAGTC	1980
Db	1921	AAGAGAAAAAATCTGAAAAATGAGAGAGGCTACTAAAGACTTTGGAACCATCAATAGTC	1980
Qy	1981	AAATGAAGAGAGCCATTGAAACAGGAGTCAAAATGTCACTTAAAGATGCGCAGAAAAAAGA	2040
Db	1981	AAATGAAGAGAGCCATTGAAACAGGAGTCAAAATGTCACTTAAAGATGCGCAGAAAAAAGA	2040
Qy	2041	TAAAAACCCACGCGCATGGAATTTGGCCAGAGCAACAGTTTAAAAAATCTCATTTATCTA	2100
Db	2041	TAAAAACCCACGCGCATGGAATTTGGCCAGAGCAACAGTTTAAAAAATCTCATTTATCTA	2100
Qy	2101	ATCAACCAAAAATTTGATGAATCTCTTTCAGTCCCAATTTGAAAAAGAGGAGTCAAAATA	2160
Db	2101	ATCAACCAAAAATTTGATGAATCTCTTTCAGTCCCAATTTGAAAAAGAGGAGTCAAAATA	2160

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 26, 2005, 14:34:00 ; Search time 1173 Seconds

(without alignments)
17403.200 Million cell updates/sec

Title: US-10-079-429A-3

Perfect score: 3063

Sequence: 1 ggcacgagtgctgttcg.....aacgtaataaaactaataac 3063

Scoring table:

Gapop 60.0 , Gapext 60.0

Searched: 4996997 seqs, 3332346308 residues

Word size : 0

Total number of hits satisfying chosen parameters: 9993994

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : N_Geneseq_21.*

- 1: Geneseqn1980s.*
- 2: Geneseqn1980s.*
- 3: Geneseqn2000s.*
- 4: Geneseqn2001as.*
- 5: Geneseqn2001bs.*
- 6: Geneseqn2002as.*
- 7: Geneseqn2002bs.*
- 8: Geneseqn2003as.*
- 9: Geneseqn2003bs.*
- 10: Geneseqn2003cs.*
- 11: Geneseqn2003ds.*
- 12: Geneseqn2004as.*
- 13: Geneseqn2004bs.*
- 14: Geneseqn2005s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3063	100.0	3063	2	Aaq97526 Human DNA
2	3063	100.0	3063	4	Aah76365 Human PMS
3	3063	100.0	3063	5	Aah75042 Nucleotid
4	3063	100.0	3063	6	Abk86089 Human PMS
5	3063	100.0	3063	6	Aad39198 Human mis
6	3063	100.0	3063	6	Aal48698 Human mis
7	3063	100.0	3063	6	Aad45354 Human MLH
8	3063	100.0	3063	6	Aad39770 Human PMS
9	3063	100.0	3063	8	Abx12940 DNA encod
10	3063	100.0	3063	8	Acab9704 cDNA enco
11	3063	100.0	3063	9	Accl3762 Human mis
12	3063	100.0	3063	9	Aal57764 Human mis
13	3063	100.0	3063	9	Ada06245 DNA encod
14	3063	100.0	3063	10	Adc89608 Human PMS
15	3063	100.0	3063	10	Ades5236 Farnesyl
16	3063	100.0	3063	10	Adf17893 Human PMS
17	3063	100.0	3063	10	Adg62892 Human PMS
18	3063	100.0	3063	10	Adh62630 Human mis
19	3063	100.0	3063	10	Adh60982 Human cDN

20	3063	100.0	3063	11	ADI32109
21	3063	100.0	3063	12	ADP78841
22	3063	100.0	3063	12	ADG46768
23	3063	100.0	3063	12	ADO40067
24	3063	100.0	3063	12	ADP66679
25	3063	100.0	3063	13	ADL3884
26	3063	100.0	3063	13	ADT98687
27	3063	100.0	3063	13	ADU77023
28	3063	100.0	3063	13	ADS84176
29	3063	100.0	3063	14	ADX58474
30	3063	100.0	3063	14	ADY53430
31	3012	98.3	3063	6	AD45433
32	3004	98.1	3239	14	AE22842
33	2879	94.0	3121	14	ADX08122
34	2110	68.9	2793	13	ACN42335
35	187	6.1	327	8	ABX12507
36	187	6.1	327	9	ADA06170
37	154	5.0	559	5	ABV48030
38	126	4.1	478	9	ACH23591
39	115	3.8	184	3	AAC22447
40	96	3.1	189	2	AAT19265
41	82	2.7	534	5	ABV18243
42	60	2.0	60	6	ABN46436
43	50	1.6	50	6	ABZ01968
44	41	1.3	493	13	ADQ79340
45	35	1.1	51	4	AAL29142

ALIGNMENTS

RESULT 1

AAQ97526

ID AAQ97526 standard; cDNA; 3063 BP.

XX

AC AAQ97526;

XX

DT 27-MAR-1996 (first entry)

XX

DE Human DNA repair protein hMLH2 coding sequence.

XX

KW DNA repair protein; hMLH1; hMLH2; hMLH3; therapy; cancer; vectors;

KW DNA synthesis; diagnosis; disease; mutL4; ds.

XX

OS Homo sapiens.

XX

FH Key Location/Qualifiers

CDS

FT 81..2873

FT /*tag= a

FT /product= "DNA repair protein."

FT /note= "The CDS may extend to position 2879 where a TGA stop codon is located. There could possibly be an error in the decoded protein in the specification since, if the stop codon is the correct translation termination signal, the C-terminal end of the protein should end with two Threonine residues. It is possible that one of these has been omitted."

W09520678-A1.

03-AUG-1995.

25-JAN-1995;

95WO-US001035.

27-JAN-1994;

94US-00187757.

16-MAR-1994;

94US-00210143.

23-AUG-1994;

94US-00294312.

(HUMA-) HUMAN GENOME SCI INC.

PA

Haseltine WA, Ruben SM, Wei Y, Adams MD, Fleischmann RD;

PI Fraser CM, Fuldner RA, Kirkness EF, Rosen CA;

XX

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 26, 2005, 14:43:11 ; Search time 7734 Seconds
(without alignments)
18529.717 Million cell updates/sec

Title: US-10-079-429A-3

Perfect score: 3063

Sequence: 1 gacacagggctgcttgcg.....aacgtaataaataataac 3063

Scoring table: OLIGO_NUC

Gapop 60.0 , Gapext 60.0

Searched: 41078325 seqs, 23393541228 residues

Word size: 0

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database: EST:*

1: gb_est1: *
2: gb_est2: *
3: gb_est3: *
4: gb_est4: *
5: gb_est5: *
6: gb_est6: *
7: gb_est7: *
8: gb_est8: *
9: gb_est9: *
10: gb_est10: *
11: gb_est11: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1804	58.9	1804	4	CR610658 full-length
2	1378	45.0	2365	4	BC036376 Homo sapi
3	1294	42.2	2667	11	DQ052952 Homo sapi
4	844	27.6	3130	4	CR859593 Pongo pyg
5	830	27.1	2667	11	DQ052953 Pan trogl
6	804	26.2	1011	5	BX353664 BX353664
7	780	25.5	780	7	CN336501 170006001
8	776	25.3	776	5	BX117693 BX117693
9	766	25.0	1021	3	BM479838 AGENCOURT
10	711	23.2	733	7	CN336498 170006001
11	705	23.0	1117	3	BM553209 AGENCOURT
12	694	22.7	722	3	BM677668 UI-E-E01-
13	677	22.1	704	5	BQ771615 UI-H-EZ1-
14	634	20.7	638	1	AL043809 DKF2P434P
15	616	20.1	898	5	BX435290 BX435290
16	615	20.1	746	5	BUE623174 BX-H-FL1-
17	604	19.7	903	5	BX327629 BX327629
18	590	19.3	641	6	CB157373 K-EST0216
19	590	19.3	748	8	CA415469 UI-H-EZ0-
20	579	18.9	668	8	DN998411 TC106042
21	574	18.7	610	1	AI636100 t929207.x
22	566	18.5	566	6	CB161021 K-EST0220

23	565	18.4	956	3	BM800196	BM800196	AGENCOURT
24	555	18.1	782	5	BQ429685	BQ429685	AGENCOURT
25	553	18.1	694	1	AL705101	AL705101	DFZP886H
26	552	18.0	603	3	BM723144	BM723144	UI-E-E01-
27	528	17.2	623	2	BE350913	BE350913	h-63h02.x
28	525	17.1	558	6	CB136057	CB136057	K-EST0188
29	525	17.1	561	5	BQ574496	BQ574496	UI-H-EZ1-
30	524	17.1	586	2	BF056020	BF056020	7k07e01.x
31	516	16.8	525	6	CB268969	CB268969	1007876.H
32	513	16.7	706	6	CB852984	CB852984	UI-CF-PNO
33	504	16.5	555	6	CB131671	CB131671	K-EST0181
34	503	16.4	893	5	BE328949	BE328949	BX328949
35	502	16.4	604	2	BE350907	BE350907	h-63g03.x
36	502	16.4	626	1	AV730735	AV730735	AV730735
37	486	15.9	500	1	AI367805	AI367805	gt56d07.x
38	486	15.9	4758	4	CR749432	CR749432	Homo sapi
39	484	15.8	534	1	AA573406	AA573406	nm53d01.8
40	484	15.8	586	7	CN336500	CN336500	170004247
41	482	15.7	538	1	AA573397	AA573397	nm53c01.8
42	481	15.7	524	1	AI660351	AI660351	wg62b03.x
43	479	15.6	492	1	AI277404	AI277404	gm58g01.x
44	477	15.6	820	5	BX419980	BX419980	EX419980
45	477	15.6	835	2	BG168340	BG168340	602342315

ALIGNMENTS

RESULT 1
CR610658
LOCUS
DEFINITION
full-length cDNA clone CS0DC010YK21 of Neuroblastoma Cot
25-normalized of Homo sapiens (human).
ACCESSION
CR610658
VERSION
CR610658.1 GI:50491465
KEYWORDS
HTC; CNSLT CDNA.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE
1 (bases 1 to 1804)
AUTHORS
Li W.B., Gruber C., Jessee J. and Polayes D.
TITLE
Full-length cDNA libraries and normalization
JOURNAL
Unpublished
REMARK
Contact : Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Faraday Avenue
Genoscope.
REFERENCE
2 (bases 1 to 1804)
Direct Submission
Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
COMMENT
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-strand cDNA was digested with Not I and cloned
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
was normalized. Library was constructed by Life Technologies, a
division of Invitrogen.
FEATURES
Location/Qualifiers
1..1804
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DC010YK21"
/tissue_type="Neuroblastoma Cot 25-normalized"
/plasmid="pCMVSPORT_6"
ORIGIN
Query Match 58.9%; Score 1804; DB 4; Length 1804;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1804; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1215 AATTATTCAAATGTTGATGACTTCAGTCATTCATTCCTCAAAATGATATGATGATGAA 1274
 1 AATTATTCAAATGTTGATGACTTCAGTCATTCATTCCTCAAAATGATATGATGATGAA 60
 1275 TCTGGAAAAACACATGATGATGTTTAAATCACCAGATAAGATTGGTGCATTTGGTTAT 1334
 61 TCTGGAAAAACACATGATGATGTTTAAATCACCAGATAAGATTGGTGCATTTGGTTAT 120
 1335 GGTCAATGTAGTAGTGAATTTCTAATCATGTATGATAAAAAACATGAAGATGCATTCAGGAC 1394
 121 GGTCAATGTAGTAGTGAATTTCTAATCATGTATGATAAAAAACATGAAGATGCATTCAGGAC 180
 1395 ATTTCAATCAGTAATCTATCATGGGAGAACTCTCAGACGGAATATATAGTAAATCTGTTTT 1454
 181 ATTTCAATCAGTAATCTATCATGGGAGAACTCTCAGACGGAATATATAGTAAATCTGTTTT 240
 1455 ATAAGTTCCTGTTAAGCACACCCAGTCAGAAAAATGGCAATAAAGACCATATAGATGAGAT 1514
 241 ATAAGTTCCTGTTAAGCACACCCAGTCAGAAAAATGGCAATAAAGACCATATAGATGAGAT 300
 1515 GGGGAAATCAGGAAGAAGCAGGTCTTGAAACTCTTCGGAATTTCTCGAGATGAGTGG 1574
 301 GGGGAAATCAGGAAGAAGCAGGTCTTGAAACTCTTCGGAATTTCTCGAGATGAGTGG 360
 1575 AGCAGGGGAATATATCTTAAATTTCAAGTGGGAGAGAAATATTGAACCTGTGAAATTTT 1634
 361 AGCAGGGGAATATATCTTAAATTTCAAGTGGGAGAGAAATATTGAACCTGTGAAATTTT 420
 1635 GTGCTGTAAGAAAGTTTACCATGTAAGTAATTAATTAATTAATTAATTAATTAATTAAT 1694
 421 GTGCTGTAAGAAAGTTTACCATGTAAGTAATTAATTAATTAATTAATTAATTAATTAAT 480
 1695 ATGAATCTTAAATGAAGATTCATGTAACAAATTAATTAATTAATTAATTAATTAATTAAT 1754
 481 ATGAATCTTAAATGAAGATTCATGTAACAAATTAATTAATTAATTAATTAATTAATTAAT 540
 1755 AAGTTTACAGTTATGATTTACTTAGCAATCGATTAATCAAGAAACCCATGTCAGCAAGT 1814
 541 AAGTTTACAGTTATGATTTACTTAGCAATCGATTAATCAAGAAACCCATGTCAGCAAGT 600
 1815 GCTCTTTTGTTCAGATCATGCTCTCAGTTTCTCATAGAAAATCCTAAGACTAGTTT 1874
 601 GCTCTTTTGTTCAGATCATGCTCTCAGTTTCTCATAGAAAATCCTAAGACTAGTTT 660
 1875 GAGGATGCAACACTACAAATTTGAAGAACTGTGGAAGACATTTAGTGAAGAGGAAAACTG 1934
 661 GAGGATGCAACACTACAAATTTGAAGAACTGTGGAAGACATTTAGTGAAGAGGAAAACTG 720
 1935 AATATGGAAGAGAGGCTACTAAGACTTTGGAACGATACAAATAGTCAAAATGGAAGAGGCC 1994
 721 AATATGGAAGAGAGGCTACTAAGACTTTGGAACGATACAAATAGTCAAAATGGAAGAGGCC 780
 1995 ATTGAACAGAGTCAAAATGTCATTAAGATGCGCAGAAAAAGATAAAACCCACGAGC 2054
 781 ATTGAACAGAGTCAAAATGTCATTAAGATGCGCAGAAAAAGATAAAACCCACGAGC 840
 2055 GCATGGAATTTGGCCCGAGAGCAAGATTAAAAACCTCATATCTAATCAACCAAACTT 2114
 841 GCATGGAATTTGGCCCGAGAGCAAGATTAAAAACCTCATATCTAATCAACCAAACTT 900
 2115 GATGAACCTCTTCAGTCCCAATTTGAAGAAAGAGGATCAAAATATTAATAATGGTACAG 2174
 901 GATGAACCTCTTCAGTCCCAATTTGAAGAAAGAGGATCAAAATATTAATAATGGTACAG 960
 2175 ATCCCTTTTCTATGAAAACTTAAAAATTAATTTAAGAAAAACAAACAAAGTTGACTTA 2234
 961 ATCCCTTTTCTATGAAAACTTAAAAATTAATTTAAGAAAAACAAACAAAGTTGACTTA 1020
 2235 GAAGAGAAGGATGAACCTCTGATCCAAATCTCAGGTTTCTGATGATGATGATGATGATG 2294
 1021 GAAGAGAAGGATGAACCTCTGATCCAAATCTCAGGTTTCTGATGATGATGATGATGATG 1080
 2295 ACATCCAAACAGAGGTAATGTTTAAATTCATATAGTAGTAGAGAGAGCCCTGCTATTT 2354

1081 ACATCCAAACAGAGGTAATGTTTAAATCCATATAGATGAGAGGCCCTGCTATTT 1140
 2355 AAAAGACTTCTTGAGAAATCATAACTTCCTGCAGAGCCACTCGGAAAAAGCCAAATTAATGTTA 2414
 1141 AAAAGACTTCTTGAGAAATCATAACTTCCTGCAGAGCCACTCGGAAAAAGCCAAATTAATGTTA 1200
 2415 ACAGAGAGTCTTTTAAATGGATCTCATATTTAGACGTTTATATATAAAATGACACAGAT 2474
 1201 ACAGAGAGTCTTTTAAATGGATCTCATATTTAGACGTTTATATATAAAATGACACAGAT 1260
 2475 GACCAAGATACAGTGGATCAACTTACCTGTCTGATCTCGTCTTTACAGCGAATGGTTTC 2534
 1261 GACCAAGATACAGTGGATCAACTTACCTGTCTGATCTCGTCTTTACAGCGAATGGTTTC 1320
 2535 AAGATAAAATGATACAGGAGTTTCAATTTACTGAAATTTACTTGGAAATAGAGAAATG 2594
 1321 AAGATAAAATGATACAGGAGTTTCAATTTACTGAAATTTACTTGGAAATAGAGAAATG 1380
 2595 GCTAATTCCTCCCAATCTATGGAGTAGCAGATTTAAAGAAATTTCTTAATGCTATATTA 2654
 1381 GCTAATTCCTCCCAATCTATGGAGTAGCAGATTTAAAGAAATTTCTTAATGCTATATTA 1440
 2655 AACAGAAATGCAAGGAAAGTTTATGAATGTAGACCTCGCAAGATGATTAAGTTATTTAGAG 2714
 1441 AACAGAAATGCAAGGAAAGTTTATGAATGTAGACCTCGCAAGATGATTAAGTTATTTAGAG 1500
 2715 GGAGAGCAGTGCCTCTATCCAGCAATTTACCCATGTAATTAAGAGAGGACATCCAA 2774
 1501 GGAGAGCAGTGCCTCTATCCAGCAATTTACCCATGTAATTAAGAGAGGACATCCAA 1560
 2775 GACATTTCTACAGAAATGAAGCAGGTTTGGAAATGAATTAAGAGAGTGTGTTCAATGT 2834
 1561 GACATTTCTACAGAAATGAAGCAGGTTTGGAAATGAATTAAGAGAGTGTGTTCAATGT 1620
 2835 CGCCCAATTTTTCATCAATTAACCTATCTTCCAGAAACTTACATGATTAATTAATGTTAAG 1680
 1621 CGCCCAATTTTTCATCAATTAACCTATCTTCCAGAAACTTACATGATTAATTAATGTTAAG 1680
 2895 AAGATTAGTACCATGAAATTTGGTCTCTCATATAAGAGAGTGTGTTTAAAT 2954
 1681 AAGATTAGTACCATGAAATTTGGTCTCTCATATAAGAGAGTGTGTTTAAAT 1740
 2955 ATCTTTGATTAATGTCACATGTTATTTTAAATGAGGATTCACATGCTGTTTAAAT 3014
 1741 ATCTTTGATTAATGTCACATGTTATTTTAAATGAGGATTCACATGCTGTTTAAAT 1800
 3015 TATT 3018
 1801 TATT 1804

RESULT 2

BC036376

LOCUS

DEFINITION

BC036376 2365 bp mRNA linear HTC 19-NOV-2003
 Homo sapiens PMS1 postmeiotic segregation increased 1 (S.
 cerevisiae), mRNA (cDNA clone IMAGE:4822417), containing
 frame-shift errors.

ACCESSION

BC036376

VERSION

BC036376.1

KEYWORDS

HTC.

SOURCE

Homo sapiens

ORGANISM

Homo sapiens (human)

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;

Hominidae; Homo.

REFERENCE

1 (bases 1 to 2365)

AUTHORS

Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G.,

Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D.,

Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K.,

Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F.,

Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,

Stapleton, M., Soares, M.B., Bonaldo, M.P., Casavant, T.L.,

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 26, 2005, 15:09:31 ; Search time 371 Seconds
(without alignments)
14675.664 Million cell updates/sec

Title: US-10-079-429A-3
Perfect score: 3063
Sequence: 1 ggcacgagggctgcttgcg.....aacgtataaataaataaac 3063

Scoring table: OLIGO NUC
Gapop 60.0 , Gapext 60.0

Searched: 1303057 seqs, 888780828 residues

Word size : 0

Total number of hits satisfying chosen parameters: 2606114

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued Patents NA:*

1: /cgn2_6/prodata/1/ina/1 COMB.seq:*

2: /cgn2_6/prodata/1/ina/5 COMB.seq:*

3: /cgn2_6/prodata/1/ina/6A COMB.seq:*

4: /cgn2_6/prodata/1/ina/6B COMB.seq:*

5: /cgn2_6/prodata/1/ina/H COMB.seq:*

6: /cgn2_6/prodata/1/ina/PCUS COMB.seq:*

7: /cgn2_6/prodata/1/ina/PP COMB.seq:*

8: /cgn2_6/prodata/1/ina/RE COMB.seq:*

9: /cgn2_6/prodata/1/ina/backfileseq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3063	100.0	3063	US-08-294-312B-3	Sequence 3, Appli
2	3063	100.0	3063	US-08-468-024B-3	Sequence 3, Appli
3	3063	100.0	3063	US-09-708-200-12	Sequence 12, Appli
4	3063	100.0	3063	US-09-023-655-1435	Sequence 1435, Ap
5	3063	100.0	3063	US-08-465-679-3	Sequence 3, Appli
6	3063	100.0	3063	US-09-788-657-8	Sequence 8, Appli
7	3063	100.0	3063	US-09-712-691-10	Sequence 10, Appli
8	3063	100.0	3063	US-09-707-468-10	Sequence 10, Appli
9	3063	100.0	3063	US-10-641-068-8	Sequence 8, Appli
10	2761	90.1	2817	US-09-949-016-1916	Sequence 1916, Ap
11	892	29.1	96845	US-09-949-016-13658	Sequence 13658, A
12	187	6.1	327	US-08-187-757D-3	Sequence 3, Appli
13	187	6.1	327	US-08-210-143C-3	Sequence 3, Appli
14	134	4.4	601	US-09-949-016-65618	Sequence 65618, A
15	115	3.8	184	US-09-513-999C-26522	Sequence 26522, A
16	50	1.6	50	US-10-131-827-1959	Sequence 1959, Ap
17	38	1.2	601	US-09-949-016-65619	Sequence 65619, A
18	23	0.8	601	US-09-949-016-151589	Sequence 151589, A
19	23	0.8	601	US-09-949-016-151591	Sequence 151591, A
20	23	0.8	143550	US-09-949-016-14143	Sequence 14143, A
21	23	0.8	169334	US-09-949-016-15999	Sequence 15999, A
22	22	0.7	60	US-08-294-312B-58	Sequence 58, Appli
23	22	0.7	60	US-08-468-024B-58	Sequence 58, Appli
24	22	0.7	60	US-08-465-679-58	Sequence 58, Appli

C 25	22	0.7	331814	3	US-09-949-016-12008	Sequence 12008, A
C 26	22	0.7	331814	3	US-09-949-016-17056	Sequence 17056, A
C 27	21	0.7	21	21	US-08-294-312B-59	Sequence 59, Appli
C 28	21	0.7	21	21	US-08-294-312B-61	Sequence 61, Appli
C 29	21	0.7	21	21	US-08-294-312B-70	Sequence 70, Appli
C 30	21	0.7	21	21	US-08-294-312B-72	Sequence 72, Appli
C 31	21	0.7	21	21	US-08-468-024B-59	Sequence 59, Appli
C 32	21	0.7	21	21	US-08-468-024B-61	Sequence 61, Appli
C 33	21	0.7	21	21	US-08-468-024B-70	Sequence 70, Appli
C 34	21	0.7	21	21	US-08-468-024B-72	Sequence 72, Appli
C 35	21	0.7	21	21	US-08-465-679-59	Sequence 59, Appli
C 36	21	0.7	21	21	US-08-465-679-61	Sequence 61, Appli
C 37	21	0.7	21	21	US-08-465-679-70	Sequence 70, Appli
C 38	21	0.7	21	21	US-08-465-679-72	Sequence 72, Appli
C 39	21	0.7	58	58	US-08-294-312B-60	Sequence 60, Appli
C 40	21	0.7	58	58	US-08-468-024B-60	Sequence 60, Appli
C 41	21	0.7	58	58	US-08-465-679-60	Sequence 60, Appli
C 42	21	0.7	601	601	US-09-949-016-60541	Sequence 60541, A
C 43	21	0.7	32099	3	US-09-949-016-16562	Sequence 16562, A
C 44	21	0.7	59519	3	US-09-949-016-13504	Sequence 13504, A
C 45	21	0.7	121970	3	US-09-949-016-17216	Sequence 17216, A

ALIGNMENTS

RESULT 1						
US-08-294-312B-3						
; Sequence 3, Application US/08294312B						
; Patent No. 6380369						
; GENERAL INFORMATION:						
; APPLICANT: Adams et al.						
; TITLE OF INVENTION: Human DNA Mismatch Repair Proteins						
; FILE REFERENCE: PF106P2						
; CURRENT APPLICATION NUMBER: US/08/294,312B						
; PRIOR FILING DATE: 1994-08-23						
; PRIOR APPLICATION NUMBER: 08/210,143						
; PRIOR FILING DATE: 1994-03-16						
; PRIOR APPLICATION NUMBER: 08/187,757						
; PRIOR FILING DATE: 1994-01-27						
; NUMBER OF SEQ ID NOS: 78						
; SOFTWARE: Patent in version 3.0						
; SEQ ID NO 3						
; LENGTH: 3063						
; TYPE: DNA						
; ORGANISM: homo sapiens						
; FEATURE:						
; NAME/KEY: CDS						
; LOCATION: (81)..(2879)						
US-08-294-312B-3						
Query Match 100.0%; Score 3063; DB 3; Length 3063;						
Best Local Similarity 100.0%; Pred. No. 0;						
Matches 3063; Conservative 0; Mismatches 0; Indels 0; Gaps 0;						
QY	1	GGCAGAGTGGCTGCTTGGCGCTAGTGGATGTAATTGCTGCTCGCTAGCAGCAAG	60			
Db	1	GGCAGAGTGGCTGCTTGGCGCTAGTGGATGTAATTGCTGCTCGCTAGCAGCAAG	60			
QY	61	CTGCTCTGTGTTAAAGCGAAATGAAACAATTCCTCGCGCAAAGTTCGACTCTTTCAA	120			
Db	61	CTGCTCTGTGTTAAAGCGAAATGAAACAATTCCTCGCGCAAAGTTCGACTCTTTCAA	120			
QY	121	GTTCCTCAGATCATCTTCGGTGGTTCAGTGGTTGTTAAAGAGCTTATTGAAAACCTCTTGG	180			
Db	121	GTTCCTCAGATCATCTTCGGTGGTTCAGTGGTTGTTAAAGAGCTTATTGAAAACCTCTTGG	180			
QY	181	ATGCTGTGTCACCAAGCGTAGATGTTAACTCGAGAACTATGAGATTGATAAATGAGG	240			
Db	181	ATGCTGTGTCACCAAGCGTAGATGTTAACTCGAGAACTATGAGATTGATAAATGAGG	240			
QY	241	TGCGAGATAACCGGGAGGGTATCAAGGGCTGTTGATGCACCTGTAATGCAATGAATCT	300			
Db	241	TGCGAGATAACCGGGAGGGTATCAAGGGCTGTTGATGCACCTGTAATGCAATGAATCT	300			

Db 241 TCGAGATTAACGGGAGGGTATCAAGGCTGTGATGCACCTGTAATGGCAATGAAGTACT 300
Qy 301 ACACCTCAAAAAATAATAGTCATGAAGATCTTGAAAAATTGACAACTTACGGTTTCGTG 360
Db 301 ACACCTCAAAAAATAATAGTCATGAAGATCTTGAAAAATTGACAACTTACGGTTTCGTG 360
Qy 361 GAGAACCTTGGGTCATTTGTTGTATAGCTGAGGTTTAAATACAAAGAACGGCTG 420
Db 361 GAGAACCTTGGGTCATTTGTTGTATAGCTGAGGTTTAAATACAAAGAACGGCTG 420
Qy 421 CTGATAATTTAGCACCCAGTATGTTTATAGTGGCAGTGGCCACATATCTTCTCAGAAAC 480
Db 421 CTGATAATTTAGCACCCAGTATGTTTATAGTGGCAGTGGCCACATATCTTCTCAGAAAC 480
Qy 481 CTTTACATCTTGTCAAGGTACAACTGTAACTGCTTTAAGATTATTTAAGAACTCTACCTG 540
Db 481 CTTTACATCTTGTCAAGGTACAACTGTAACTGCTTTAAGATTATTTAAGAACTCTACCTG 540
Qy 541 TAAGAAAGCAGTTTACTCACTGCAAAAAAATGTAAGATGAATAAAAAAGATCCAAG 600
Db 541 TAAGAAAGCAGTTTACTCACTGCAAAAAAATGTAAGATGAATAAAAAAGATCCAAG 600
Qy 601 ATCTCTCATGAGCTTGGTATCCTTAAACCTGACTTAAGCATTTGCTTTTGTACATAACA 660
Db 601 ATCTCTCATGAGCTTGGTATCCTTAAACCTGACTTAAGCATTTGCTTTTGTACATAACA 660
Qy 661 AGGCAGTTAATGGCAGAAAAAGCAGATATCAGATCAAAAGATGGCTCTCATGTCAGTTC 720
Db 661 AGGCAGTTAATGGCAGAAAAAGCAGATATCAGATCAAAAGATGGCTCTCATGTCAGTTC 720
Qy 721 TGGGGACTGCTGTATGAACAAATATGAATCCCTTTCAGTACCCTCTGAAAGATCTCAGA 780
Db 721 TGGGGACTGCTGTATGAACAAATATGAATCCCTTTCAGTACCCTCTGAAAGATCTCAGA 780
Qy 781 TTTTATCTCAGTGGATTTCTTCAAAGTGTGATGCAGACCACTCTTTTCACTAGTCTTTCAA 840
Db 781 TTTTATCTCAGTGGATTTCTTCAAAGTGTGATGCAGACCACTCTTTTCACTAGTCTTTCAA 840
Qy 841 CACCAGAAAGAGTTTCATCTTCAATAACAGTGCAGCAGTACATCAAAAAGATATCTTAA 900
Db 841 CACCAGAAAGAGTTTCATCTTCAATAACAGTGCAGCAGTACATCAAAAAGATATCTTAA 900
Qy 901 AGTTAATCCGACATCATTAACCTGAATGCCTTAAGGAATCTACTCGTTTGTATCCTG 960
Db 901 AGTTAATCCGACATCATTAACCTGAATGCCTTAAGGAATCTACTCGTTTGTATCCTG 960
Qy 961 TTTTCTTCTGAAATCGATGTTCTTACAGTGTGATGTTGATGTAATTTAACCACAGATA 1020
Db 961 TTTTCTTCTGAAATCGATGTTCTTACAGTGTGATGTTGATGTAATTTAACCACAGATA 1020
Qy 1021 AAAGCCAAAGTATTATTAACAAATAAGGAATCTGTTTAAATTTGCTTTGAAATCTGATGA 1080
Db 1021 AAAGCCAAAGTATTATTAACAAATAAGGAATCTGTTTAAATTTGCTTTGAAATCTGATGA 1080
Qy 1081 CGACTTGTATGACCAATTAAGTACAAATCTTATGAAATTAATAAACAAGATGTTT 1140
Db 1081 CGACTTGTATGACCAATTAAGTACAAATCTTATGAAATTAATAAACAAGATGTTT 1140
Qy 1141 CCGCAGCTGACATCGTTCTTGTATGTAACACAGAGAAACAGATGTCGTTTTTAAATAAGTGG 1200
Db 1141 CCGCAGCTGACATCGTTCTTGTATGTAACACAGAGAAACAGATGTCGTTTTTAAATAAGTGG 1200
Qy 1201 AATCATCTGGAAGAAATTTATCAAAATGTTGATATCTTCAGTCATTTCCATTTCCAAATGATA 1260
Db 1201 AATCATCTGGAAGAAATTTATCAAAATGTTGATATCTTCAGTCATTTCCATTTCCAAATGATA 1260
Qy 1261 TGCATAATGATGAATCTGGAAGAAACACATGATGATTTTAAATCCACAGATAGTATG 1320
Db 1261 TGCATAATGATGAATCTGGAAGAAACACATGATGATTTTAAATCCACAGATAGTATG 1320
Qy 1321 GTGACTTTTGGTTATGGTCATTTAGTATGTAATTTCTTAACTTGAATAAAAAACACTAAGA 1380
Db 1321 GTGACTTTTGGTTATGGTCATTTAGTATGTAATTTCTTAACTTGAATAAAAAACACTAAGA 1380

Qy 1381 ATGCATTTTCAGGACATTTCAATGAGTAAATGTATCATGGAGAACTCTCAGACGGAATATA 1440
Db 1381 ATGCATTTTCAGGACATTTCAATGAGTAAATGTATCATGGAGAACTCTCAGACGGAATATA 1440
Qy 1441 GTAAAACTTGTTTTATAAGTTCCGTTAAGCACACCCAGTCAGAAAAATGGCAATAAGACC 1500
Db 1441 GTAAAACTTGTTTTATAAGTTCCGTTAAGCACACCCAGTCAGAAAAATGGCAATAAGACC 1500
Qy 1501 ATATAGATCAGAGTGGGGAATATGAGGAAGAGCAGGCTCTTCAAACTCTTCGGAATTTT 1560
Db 1501 ATATAGATCAGAGTGGGGAATATGAGGAAGAGCAGGCTCTTCAAACTCTTCGGAATTTT 1560
Qy 1561 CTGTCAGATCAGTGGGAGCGGGAATATATCTTAAAAATTCAGTGGGAGAGAAATATTTGAAC 1620
Db 1561 CTGTCAGATCAGTGGGAGCGGGAATATATCTTAAAAATTCAGTGGGAGAGAAATATTTGAAC 1620
Qy 1621 CTGTGAAAAATTTTGTGCTGAAAAAGTTTACCATGTAAAGTAAAGTAAATATATATTC 1680
Db 1621 CTGTGAAAAATTTTGTGCTGAAAAAGTTTACCATGTAAAGTAAAGTAAATATATATTC 1680
Qy 1681 CAATCCCTGAACAAATGATCTTAAATGAAGATTTCATGTAAACAAAAATCAAAATGTATATAG 1740
Db 1681 CAATCCCTGAACAAATGATCTTAAATGAAGATTTCATGTAAACAAAAATCAAAATGTATATAG 1740
Qy 1741 AATAATAAATCTGAAAAAGTTACAGCTTATGATTTACTTAGCAATCGAGTAATCAAGAAAC 1800
Db 1741 AATAATAAATCTGAAAAAGTTACAGCTTATGATTTACTTAGCAATCGAGTAATCAAGAAAC 1800
Qy 1801 CCATGTCAGCAAGTGTCTTTTGTGTTTCAAGATCATGCTCCTCAGTTTCTCATAGAAAAATC 1860
Db 1801 CCATGTCAGCAAGTGTCTTTTGTGTTTCAAGATCATGCTCCTCAGTTTCTCATAGAAAAATC 1860
Qy 1861 CTAAGACTAGTTTACAGGATGCAACACTACAAATTTGAAGAACTGTGGAAGACATTCAGTG 1920
Db 1861 CTAAGACTAGTTTACAGGATGCAACACTACAAATTTGAAGAACTGTGGAAGACATTCAGTG 1920
Qy 1921 AAGAGGAAAACTGAAAAATATGAAGAGAGGCTACTAAAGACTTGGAAACGATACAATAGTC 1980
Db 1921 AAGAGGAAAACTGAAAAATATGAAGAGAGGCTACTAAAGACTTGGAAACGATACAATAGTC 1980
Qy 1981 AAATGAAGAGAGCCATTGAAAGAGGATCAAAATGTCACTAAAAGATGGCAGAGAAAAAGA 2040
Db 1981 AAATGAAGAGAGCCATTGAAAGAGGATCAAAATGTCACTAAAAGATGGCAGAGAAAAAGA 2040
Qy 2041 TAAACCCACCGCATCGGAATTTGGCCAGAGCACAAGTTAAAAAACCCTCATTTATCTA 2100
Db 2041 TAAACCCACCGCATCGGAATTTGGCCAGAGCACAAGTTAAAAAACCCTCATTTATCTA 2100
Qy 2101 ATCAACCAAAAACCTTGATGAACCTCTTCAGTCCCAAAATGGAAGAGAGGAGTCAAAAATA 2160
Db 2101 ATCAACCAAAAACCTTGATGAACCTCTTCAGTCCCAAAATGGAAGAGAGGAGTCAAAAATA 2160
Qy 2161 TTAATAATGGTACAGATCCCTTTTCTATGAAAACTTAAATAAATTTTAAAGAAACAAA 2220
Db 2161 TTAATAATGGTACAGATCCCTTTTCTATGAAAACTTAAATAAATTTTAAAGAAACAAA 2220
Qy 2221 ACAAGCTGCACTTAGAAGAGAGAGGATGAACCTTGTGATCCCAATCTCAGGTTTCTTG 2280
Db 2221 ACAAGCTGCACTTAGAAGAGAGAGGATGAACCTTGTGATCCCAATCTCAGGTTTCTTG 2280
Qy 2281 ATGCATGGCTTAATGACATCCAAACAGAGGATATGTTATTAATCCATATAGAGTAGAAG 2340
Db 2281 ATGCATGGCTTAATGACATCCAAACAGAGGATATGTTATTAATCCATATAGAGTAGAAG 2340
Qy 2341 AAGCCCTGCTATTTAAAGACTCTTGGAGAACTTCTGAGAGCCACTGGA 2400
Db 2341 AAGCCCTGCTATTTAAAGACTCTTGGAGAACTTCTGAGAGCCACTGGA 2400
Qy 2401 AGCCATATATGTTAAACAGAGAGTCTTTTAAATCGATCTCATTTATTTAGAGCTTTTATATA 2460
Db 2401 AGCCATATATGTTAAACAGAGAGTCTTTTAAATCGATCTCATTTATTTAGAGCTTTTATATA 2460

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 26, 2005, 18:21:20 ; Search time 1567 Seconds
(without alignments)
16164.075 Million cell updates/sec

Title: US-10-079-429A-3
Perfect score: 3063
Sequence: 1 ggcacgagtggtcgcttgcg.....aacgtaataaactaataac 3063

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 9793542 seqs, 4134689005 residues

Word size : 0

Total number of hits satisfying chosen parameters: 19587084

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Published Applications NA_Main:
1: /cgn2_6/prodata/1/pubpna/US07_PUBCOMB.seq:
2: /cgn2_6/prodata/1/pubpna/US08_PUBCOMB.seq:
3: /cgn2_6/prodata/1/pubpna/US09A_PUBCOMB.seq:
4: /cgn2_6/prodata/1/pubpna/US09B_PUBCOMB.seq:
5: /cgn2_6/prodata/1/pubpna/US10A_PUBCOMB.seq:
6: /cgn2_6/prodata/1/pubpna/US10B_PUBCOMB.seq:
7: /cgn2_6/prodata/1/pubpna/US10C_PUBCOMB.seq:
8: /cgn2_6/prodata/1/pubpna/US10D_PUBCOMB.seq:
9: /cgn2_6/prodata/1/pubpna/US10E_PUBCOMB.seq:
10: /cgn2_6/prodata/1/pubpna/US11_PUBCOMB.seq:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3063	100.0	3063	3	US-09-788-657-8 Sequence 8, Appl1
2	3063	100.0	3063	3	US-09-912-697-7 Sequence 7, Appl1
3	3063	100.0	3063	3	US-09-760-285-19 Sequence 19, Appl1
4	3063	100.0	3063	5	US-10-079-429-3 Sequence 3, Appl1
5	3063	100.0	3063	6	US-10-270-839-28 Sequence 28, Appl1
6	3063	100.0	3063	6	US-10-243-130-10 Sequence 10, Appl1
7	3063	100.0	3063	6	US-10-371-857-14 Sequence 14, Appl1
8	3063	100.0	3063	6	US-10-371-634-8 Sequence 8, Appl1
9	3063	100.0	3063	6	US-10-348-074-6 Sequence 6, Appl1
10	3063	100.0	3063	6	US-10-369-845-12 Sequence 12, Appl1
11	3063	100.0	3063	7	US-10-641-643-1435 Sequence 1435, Ap
12	3063	100.0	3063	7	US-10-641-068-8 Sequence 8, Appl1
13	3063	100.0	3063	7	US-10-283-975A-453 Sequence 453, App
14	3063	100.0	3063	7	US-10-813-502-10 Sequence 10, Appl
15	3063	100.0	3063	8	US-10-714-228-1 Sequence 1, Appl1
16	3063	100.0	3063	8	US-10-850-370-10 Sequence 10, Appl
17	3063	100.0	3063	9	US-10-933-034-1 Sequence 1, Appl1
18	3063	100.0	3063	9	US-10-901-650-10 Sequence 10, Appl
19	3063	100.0	3063	10	US-11-056-776-14 Sequence 14, Appl
20	2799	91.4	2799	6	US-10-109-791A-9 Sequence 9, Appl1
21	560	18.3	611	4	US-09-925-065A-562210 Sequence 562210,
22	233	7.6	419	7	US-10-242-535A-25367 Sequence 25367, A
23	233	7.6	419	7	US-10-085-783A-25367 Sequence 25367, A

24	154	5.0	559	8	US-10-357-930-48049	Sequence 48049, A
25	126	4.1	478	3	US-09-918-995-10803	Sequence 10803, A
26	82	2.7	534	3	US-10-357-930-18234	Sequence 18234, A
27	60	2.0	60	3	US-09-908-975-19184	Sequence 19184, A
28	50	1.6	50	6	US-10-131-827-1959	Sequence 1959, Ap
29	48	1.6	120	7	US-10-242-535A-21302	Sequence 21302, A
30	48	1.6	120	7	US-10-085-783A-21302	Sequence 21302, A
31	48	1.6	571	4	US-09-925-065A-562206	Sequence 562206,
32	48	1.6	622	4	US-09-925-065A-562207	Sequence 562207,
33	48	1.6	622	4	US-09-925-065A-562208	Sequence 562208,
34	25	0.8	25	9	US-10-792-374-6	Sequence 6, Appli
35	25	0.8	25	10	US-11-036-317-87094	Sequence 87094, A
36	25	0.8	25	10	US-11-036-317-118726	Sequence 118726,
37	25	0.8	25	10	US-11-036-317-175306	Sequence 175306,
38	25	0.8	25	10	US-11-036-317-205992	Sequence 205992,
39	25	0.8	25	10	US-11-036-317-222017	Sequence 222017,
40	25	0.8	25	10	US-11-036-317-233940	Sequence 233940,
41	25	0.8	25	10	US-11-036-317-281574	Sequence 281574,
42	25	0.8	25	10	US-11-036-317-368945	Sequence 368945,
43	23	0.8	478	7	US-10-702-075-254	Sequence 254, App
44	22	0.7	60	5	US-10-079-429-58	Sequence 58, Appl
45	22	0.7	575	4	US-09-925-065A-272278	Sequence 272278,

ALIGNMENTS

RESULT 1

US-09-788-657-8
; Sequence 8, Application US/09788657
; Patent No. US20020123149A1
; GENERAL INFORMATION:
; APPLICANT: Nicolaides, Nicholas
; APPLICANT: Sasse, Philip
; APPLICANT: Kinzler, Kenneth
; APPLICANT: Grasso, Luigi
; APPLICANT: Vogelstein, Bert
; TITLE OF INVENTION: Methods for generating hypermutable
; TITLE OF INVENTION: Yeast
; FILE REFERENCE: 01107.00097
; CURRENT APPLICATION NUMBER: US/09/788,657
; CURRENT FILING DATE: 2001-02-21
; PRIOR APPLICATION NUMBER: 60/184,336
; PRIOR FILING DATE: 2000-02-23
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 8
; LENGTH: 3063
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-788-657-8

Query Match	100.0%;	Score 3063;	DB 3;	Length 3063;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 3063;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	GGCACAGTGGCTTGGCTGGCTAGTGGATGTAATTCCTGCTCGCTAGCAGCAAG	60	
Db	1	GGCACAGTGGCTTGGCTGGCTAGTGGATGTAATTCCTGCTCGCTAGCAGCAAG	60	
QY	61	CTGCTCTGTTAAAGCGAAATGAACAATTCCTGCGGCAACAGTTCGACTCCTTTCAA	120	
Db	61	CTGCTCTGTTAAAGCGAAATGAACAATTCCTGCGGCAACAGTTCGACTCCTTTCAA	120	
QY	121	GTCTCAGATCATCACTTCGGTGGTGGTGTGTTGTAAGAGCTTATGAAAACTCTCTGG	180	
Db	121	GTCTCAGATCATCACTTCGGTGGTGGTGTGTTGTAAGAGCTTATGAAAACTCTCTGG	180	
QY	181	ATGCTGGTCCACACCGCTAGATGTTAACTGGAGACTATGATTTGATAAATTCAGG	240	
Db	181	ATGCTGGTCCACACCGCTAGATGTTAACTGGAGACTATGATTTGATAAATTCAGG	240	
QY	241	TGCGAGATAACGGGGAGGGGTATCAAGGCTGTTGATGCACCTGTAATGGCAATGAAGTACT	300	

Db 241 TCGGAGATAAACGGGAGGGTATCAAGGCTGTTGATGCACTGTAATGGCAATGAAGTACT 300
Qy
Db 301 ACACCTCAAAAATAATAGTCATGAAGATCTTGAAAATTTGACAACTTACGGTTTTCGTG 360
Qy
Db 301 ACACCTCAAAAATAATAGTCATGAAGATCTTGAAAATTTGACAACTTACGGTTTTCGTG 360
Qy
Db 361 GAGAGCCCTGGGGTCAATTTGTTGATAGCTAGAGGTTTAAATTACAAAGAAAGGGCTG 420
Qy
Db 361 GAGAGCCCTGGGGTCAATTTGTTGATAGCTAGAGGTTTAAATTACAAAGAAAGGGCTG 420
Qy
Db 421 CTGATTAATTTAGCAACCCAGTATGTTTATAGTGGCAGTGCACATACCTTTCTCGAAGAC 480
Qy
Db 421 CTGATTAATTTAGCAACCCAGTATGTTTATAGTGGCAGTGCACATACCTTTCTCGAAGAC 480
Qy
Db 481 CTTTCATCTTGCTCAAGGTCAAACTGTAACTGCTTTTAAGATTTATTAAGAAATCTACCTG 540
Qy
Db 481 CTTTCATCTTGCTCAAGGTCAAACTGTAACTGCTTTTAAGATTTATTAAGAAATCTACCTG 540
Qy
Db 541 TAAGAAAGCAGTTTTTACTCAACTGCAAAAATAATGTTAAAGATGAATAAAAAAGATCCAG 600
Qy
Db 541 TAAGAAAGCAGTTTTTACTCAACTGCAAAAATAATGTTAAAGATGAATAAAAAAGATCCAG 600
Qy
Db 601 ATCTCTCATGAGCTTTGCTTACCTTAAACCTGACCTTAAAGGATGCTTTGTACATAACA 660
Qy
Db 601 ATCTCTCATGAGCTTTGCTTACCTTAAACCTGACCTTAAAGGATGCTTTGTACATAACA 660
Qy
Db 661 AGGCACTTTATTTGGCAGAAAAGCAGAGTATCAGATCACAAGATGGCTCTCATGTCAAGTTC 720
Qy
Db 661 AGGCACTTTATTTGGCAGAAAAGCAGAGTATCAGATCACAAGATGGCTCTCATGTCAAGTTC 720
Qy
Db 721 TGGGGACTGCTGTTATGAACAATATGGAATCTCTTTCAAGTACCACTCTGAAGAAATCTCAGA 780
Qy
Db 721 TGGGGACTGCTGTTATGAACAATATGGAATCTCTTTCAAGTACCACTCTGAAGAAATCTCAGA 780
Qy
Db 781 TTTATCTCAGTGGATTTCTTCCAAAGTGTGATGCAGACCACTCTTTCACTAGTCTTTCAA 840
Qy
Db 781 TTTATCTCAGTGGATTTCTTCCAAAGTGTGATGCAGACCACTCTTTCACTAGTCTTTCAA 840
Qy
Db 841 CACCAGAAAGAGTTTTCATCTTCAATAACAGTGCAGCCAGTACATCAAAAAGATATCTTAA 900
Qy
Db 841 CACCAGAAAGAGTTTTCATCTTCAATAACAGTGCAGCCAGTACATCAAAAAGATATCTTAA 900
Qy
Db 901 AGTTAATCCGACATCAATTAACAATCTGAAATGCTTAAAGGAATCTACTCGTTTGTATCTTG 960
Qy
Db 901 AGTTAATCCGACATCAATTAACAATCTGAAATGCTTAAAGGAATCTACTCGTTTGTATCTTG 960
Qy
Db 961 TTTTCTTTCTGAAAATCGATGTTCTTACAGCTGATGTTGATGTAAATTTTAAACACCAAGATA 1020
Qy
Db 961 TTTTCTTTCTGAAAATCGATGTTCTTACAGCTGATGTTGATGTAAATTTTAAACACCAAGATA 1020
Qy
Db 1021 AAAGCCCAAGTATTATTAACAAAATAAGGAATCTGTTTAAATTTGCTCTTTGAAAATCTGATGA 1080
Qy
Db 1021 AAAGCCCAAGTATTATTAACAAAATAAGGAATCTGTTTAAATTTGCTCTTTGAAAATCTGATGA 1080
Qy
Db 1081 CGACTGTTATGACCAATTAACCTAGTACAAATCTTATGAAAATTAATAAAGATGTTT 1140
Qy
Db 1081 CGACTGTTATGACCAATTAACCTAGTACAAATCTTATGAAAATTAATAAAGATGTTT 1140
Qy
Db 1141 CCGCAGCTGATCGTTCTTTAGTAAACACAGCAAGACAGATGCTTTTAAATAAAGTGG 1200
Qy
Db 1141 CCGCAGCTGATCGTTCTTTAGTAAACACAGCAAGACAGATGCTTTTAAATAAAGTGG 1200
Qy
Db 1201 AATCATCTGAAAAGAAATTTTCAAAATGTTGATATCTTCACTTCACTTCCAAATGATA 1260
Qy
Db 1201 AATCATCTGAAAAGAAATTTTCAAAATGTTGATATCTTCACTTCACTTCCAAATGATA 1260
Qy
Db 1261 TGCATAATGATGAATCTGGAAAAAACAACCTGATGATGTTTAAATCACCAGATTAAGTATTG 1320
Qy
Db 1261 TGCATAATGATGAATCTGGAAAAAACAACCTGATGATGTTTAAATCACCAGATTAAGTATTG 1320
Qy
Db 1321 GTCACTTGGTTATGGTCATTTAGTAGTGAATTTCTTCACTTGAATTAATAAACAACCTAAGA 1380

Db 1321 GTGACTTTGGTTATGGTCAATTTAGTAGTGAATTTCTTAAACATTTGATAAAAACACTAAGA 1380
Qy
Db 1381 ATGCATTTCAAGACAAATTTCAATGAGTAATGATCATGGAGAACTCTCAGACGGAATATA 1440
Qy
Db 1381 ATGCATTTCAAGACAAATTTCAATGAGTAATGATCATGGAGAACTCTCAGACGGAATATA 1440
Qy
Db 1441 GTAAAACTGTTTATATAAGTTCCGTTAAAGCACACCAGTCAAGAAATGGCAATAAAGACC 1500
Qy
Db 1441 GTAAAACTGTTTATATAAGTTCCGTTAAAGCACACCAGTCAAGAAATGGCAATAAAGACC 1500
Qy
Db 1501 ATATAGATGAGAGTGGGAAAATGAGGAAGAGAGGCTCTTGAATACTCTTCGGAAATTT 1560
Qy
Db 1501 ATATAGATGAGAGTGGGAAAATGAGGAAGAGAGGCTCTTGAATACTCTTCGGAAATTT 1560
Qy
Db 1561 CTGCAAGATGAGTGGAGCAGGCGAAATATACCTTAAATAATTCAGTGGGAGAGAAATTTGAAC 1620
Qy
Db 1561 CTGCAAGATGAGTGGAGCAGGCGAAATATACCTTAAATAATTCAGTGGGAGAGAAATTTGAAC 1620
Qy
Db 1621 CTGTGAAAATTTTGTAGTGCCTGAAAAAAGTTTACCAATGTAAGTAAGTAATAATAATATTC 1680
Qy
Db 1621 CTGTGAAAATTTTGTAGTGCCTGAAAAAAGTTTACCAATGTAAGTAAGTAATAATAATATTC 1680
Qy
Db 1681 CAATCCCTGAAACAAATGAATCTTAAATGAAGATTCATGTAAACAAAATCAAAATGTAATAG 1740
Qy
Db 1681 CAATCCCTGAAACAAATGAATCTTAAATGAAGATTCATGTAAACAAAATCAAAATGTAATAG 1740
Qy
Db 1741 ATAATAAATCTGAAAAGTTACAGCTTATGATTTACTTAGCAATCGAGTAATCAAGAAAC 1800
Qy
Db 1741 ATAATAAATCTGAAAAGTTACAGCTTATGATTTACTTAGCAATCGAGTAATCAAGAAAC 1800
Qy
Db 1801 CCAATGTCAAGAGTCTCTTTTGTTCAGAGTCACTCGTCTCAGTTTCTCATAGAAAATTC 1860
Qy
Db 1801 CCAATGTCAAGAGTCTCTTTTGTTCAGAGTCACTCGTCTCAGTTTCTCATAGAAAATTC 1860
Qy
Db 1861 CTAAGACTAGTTTAGAGGATGCAACACTACAAATTTGAAGAACTGTGGAAGACATTCAGTG 1920
Qy
Db 1861 CTAAGACTAGTTTAGAGGATGCAACACTACAAATTTGAAGAACTGTGGAAGACATTCAGTG 1920
Qy
Db 1921 AAGAGGAAAACCTGAAATATGAAGAGAGGCTTAAAGACTTGGAAACGATCAATATAGTTC 1980
Qy
Db 1921 AAGAGGAAAACCTGAAATATGAAGAGAGGCTTAAAGACTTGGAAACGATCAATATAGTTC 1980
Qy
Db 1981 AAATGAAGAGAGCCATTGAACAGGAGTCAAAATGTCACTAAAAGATGGCAGAAAAAGA 2040
Qy
Db 1981 AAATGAAGAGAGCCATTGAACAGGAGTCAAAATGTCACTAAAAGATGGCAGAAAAAGA 2040
Qy
Db 2041 TAAAAACCCACAGCGCATGGAAATTTGGCCCGCAGAGCACAAGTTTAAAAACCTCATTTATCTA 2100
Qy
Db 2041 TAAAAACCCACAGCGCATGGAAATTTGGCCCGCAGAGCACAAGTTTAAAAACCTCATTTATCTA 2100
Qy
Db 2101 ATCAACCAAACTTGAAGAGAGGATGAACTTGTGTTGATCCCAATCTCAGGTTTCCCTG 2160
Qy
Db 2101 ATCAACCAAACTTGAAGAGAGGATGAACTTGTGTTGATCCCAATCTCAGGTTTCCCTG 2160
Qy
Db 2161 TTAATAATGTTACAGATCCCTTTTCTATGAAAACCTTAAATAATAATTTTAAAGAAACAAA 2220
Qy
Db 2161 TTAATAATGTTACAGATCCCTTTTCTATGAAAACCTTAAATAATAATTTTAAAGAAACAAA 2220
Qy
Db 2221 ACAAAGTTCATTTAGAGAGAGGATGAACTTGTGTTGATCCCAATCTCAGGTTTCCCTG 2280
Qy
Db 2221 ACAAAGTTCATTTAGAGAGAGGATGAACTTGTGTTGATCCCAATCTCAGGTTTCCCTG 2280
Qy
Db 2281 ATGCATGGCTTAATGATCCAAACACAGAGGTAATGTTTAAATCCATATAGAGTAGAAG 2340
Qy
Db 2281 ATGCATGGCTTAATGATCCAAACACAGAGGTAATGTTTAAATCCATATAGAGTAGAAG 2340
Qy
Db 2341 AAGCCCTGCTATTTAAAGACTTCTTGAAGATCATAACTCTCTGAGAGCCACTGGAAA 2400
Qy
Db 2341 AAGCCCTGCTATTTAAAGACTTCTTGAAGATCATAACTCTCTGAGAGCCACTGGAAA 2400
Qy
Db 2401 AGCCAAATATGTTAAACAGAGAGTCTTTTAAATGGAATCTCATTTATTTAGAGCTTTTATATA 2460
Qy
Db 2401 AGCCAAATATGTTAAACAGAGAGTCTTTTAAATGGAATCTCATTTATTTAGAGCTTTTATATA 2460

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 26, 2005, 20:30:17 ; Search time 244 Seconds
(without alignments)
6514.585 Million cell updates/sec

Title: US-10-079-429A-3
Perfect score: 3063
Sequence: 1 ggcacgagtggtgctgctg.....aacgtaataaactaataac 3063

Scoring table: OLIGO_NUC

Gapop 60.0 , Gapext 60.0

Searched: 4168288 seqs, 259477437 residues

Word size : 10

Total number of hits satisfying chosen parameters: 8336576

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Published Applications NA New:

- 1: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq.*
- 2: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq.*
- 3: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq.*
- 4: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq.*
- 5: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq.*
- 6: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq.*
- 7: /cgn2_6/ptodata/1/pubpna/US12_NEW_PUB.seq.*
- 8: /cgn2_6/ptodata/1/pubpna/US13_NEW_PUB.seq.*
- 9: /cgn2_6/ptodata/1/pubpna/US14_NEW_PUB.seq.*
- 10: /cgn2_6/ptodata/1/pubpna/US15_NEW_PUB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3063	100.0	3063	US-11-188-743-8	Sequence 8, Appli
2	25	0.8	25	US-11-121-849-235011	Sequence 235011,
3	25	0.8	25	US-11-121-849-235012	Sequence 235012,
4	25	0.8	25	US-11-121-849-235013	Sequence 235013,
5	25	0.8	25	US-11-121-849-235014	Sequence 235014,
6	25	0.8	25	US-11-121-849-235015	Sequence 235015,
7	25	0.8	25	US-11-121-849-235016	Sequence 235016,
8	25	0.8	25	US-11-121-849-235017	Sequence 235017,
9	25	0.8	25	US-11-121-849-235018	Sequence 235018,
10	25	0.8	25	US-11-121-849-235019	Sequence 235019,
11	25	0.8	25	US-11-121-849-235020	Sequence 235020,
12	25	0.8	25	US-11-121-849-235021	Sequence 235021,
13	21	0.7	2007	US-10-750-185-28749	Sequence 28749, A
14	20	0.7	201	US-10-995-561-78264	Sequence 78264, A
15	20	0.7	201	US-10-995-561-78285	Sequence 78285, A
16	20	0.7	201	US-10-995-561-78344	Sequence 78344, A
17	20	0.7	201	US-10-995-561-78367	Sequence 78367, A
18	20	0.7	1593	US-10-750-185-31134	Sequence 31134, A
19	20	0.7	3238	US-10-750-185-53560	Sequence 53560, A
20	20	0.7	53331	US-10-995-561-13476	Sequence 13476, A
21	20	0.7	177175	US-11-121-086-79	Sequence 79, Appl
22	19	0.6	19	US-11-101-244-382080	Sequence 382080,
23	19	0.6	19	US-11-101-244-382081	Sequence 382081,

24	19	0.6	19	US-11-101-244-382082	Sequence 382082,
25	19	0.6	19	US-11-101-244-382083	Sequence 382083,
26	19	0.6	19	US-11-101-244-382084	Sequence 382084,
27	19	0.6	19	US-11-101-244-382085	Sequence 382085,
28	19	0.6	19	US-11-101-244-382086	Sequence 382086,
29	19	0.6	19	US-11-101-244-382087	Sequence 382087,
30	19	0.6	19	US-11-101-244-382088	Sequence 382088,
31	19	0.6	19	US-11-101-244-382089	Sequence 382089,
32	19	0.6	19	US-11-101-244-382090	Sequence 382090,
33	19	0.6	19	US-11-101-244-382091	Sequence 382091,
34	19	0.6	19	US-11-101-244-382092	Sequence 382092,
35	19	0.6	19	US-11-101-244-382093	Sequence 382093,
36	19	0.6	19	US-11-101-244-382094	Sequence 382094,
37	19	0.6	19	US-11-101-244-382095	Sequence 382095,
38	19	0.6	19	US-11-101-244-382096	Sequence 382096,
39	19	0.6	19	US-11-101-244-382097	Sequence 382097,
40	19	0.6	19	US-11-101-244-382098	Sequence 382098,
41	19	0.6	19	US-11-101-244-382099	Sequence 382099,
42	19	0.6	19	US-11-101-244-382100	Sequence 382100,
43	19	0.6	19	US-11-101-244-382101	Sequence 382101,
44	19	0.6	19	US-11-101-244-382102	Sequence 382102,
45	19	0.6	19	US-11-101-244-382103	Sequence 382103,

ALIGNMENTS

RESULT 1
US-11-188-743-8
; Sequence 8, Application US/11188743
; Publication No. US20050272140A1
; GENERAL INFORMATION:
; APPLICANT: Nicolaides, Nicholas
; APPLICANT: Sasse, Philip
; APPLICANT: Kinzler, Kenneth
; APPLICANT: Grasso, Luigi
; APPLICANT: Vogelstein, Bert
; TITLE OF INVENTION: Methods for generating hypermutable
; FILE REFERENCE: 01107.00097
; CURRENT APPLICATION NUMBER: US/11/188,743
; CURRENT FILING DATE: 2005-07-26
; PRIOR APPLICATION NUMBER: US/10/641,068
; PRIOR FILING DATE: 2003-08-15
; PRIOR APPLICATION NUMBER: US/09/788,657
; PRIOR FILING DATE: 2001-02-21
; PRIOR APPLICATION NUMBER: 60/184,336
; PRIOR FILING DATE: 2000-02-23
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 8
; LENGTH: 3063
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-188-743-8

Query Match 100.0%; Score 3063; DB 7; Length 3063;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 3063; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	GGCAGAGTGCTGCTGGCGCTAGTGGATGTAATTCCTCGCTCGCGCTAGCAGCAAG	60
Db	1	GGCAGAGTGCTGCTGGCGCTAGTGGATGTAATTCCTCGCTCGCGCTAGCAGCAAG	60
QY	61	CTGCTCTGTTAAAGCGAAATGAAACAAATTCCTCGCGCAACAGTTTCGACTCTTTCAA	120
Db	61	CTGCTCTGTTAAAGCGAAATGAAACAAATTCCTCGCGCAACAGTTTCGACTCTTTCAA	120
QY	121	GTTCTCAGATCATCATTTCGCTGCTAGTGTGTTAAAGAGCTTATTGAAACCTCTTGG	180
Db	121	GTTCTCAGATCATCATTTCGCTGCTAGTGTGTTAAAGAGCTTATTGAAACCTCTTGG	180
QY	181	ATGCTGTGTCACACGCGTAGATGTTAACTGGAGAACTATGGATTTTGATAAATTTGAGG	240

181 ATGCTGGTCCCAAGCGTAGATGTTAACTGGAGAACTATGGATTTGATAAAATTTGAGG 240
241 TCGGAGATAACGGGAGGGGTATCAAGGCTGTTGATGCACCTGTAATGGCAATGAAGTACT 300
241 TCGGAGATAACGGGAGGGGTATCAAGGCTGTTGATGCACCTGTAATGGCAATGAAGTACT 300
301 ACACCTCAAAAATAAATAGTCAATGAAGATCTTGAAATTTGACAACTTACGGTTTTCGTTG 360
301 ACACCTCAAAAATAAATAGTCAATGAAGATCTTGAAATTTGACAACTTACGGTTTTCGTTG 360
361 GAGAGCCTTGGGGTCAATTTGTTGATAGCTGAGGTTTAAATTAACAACAAGAAAGCGCTG 420
361 GAGAGCCTTGGGGTCAATTTGTTGATAGCTGAGGTTTAAATTAACAACAAGAAAGCGCTG 420
421 CTGATAATTTTACACCCAGTATGTTTATAGTGGCAGTGGCCACATATCTTTCTCAAAAAC 480
421 CTGATAATTTTACACCCAGTATGTTTATAGTGGCAGTGGCCACATATCTTTCTCAAAAAC 480
481 CTTCAATCTTGGTCAAGGTACAACTGTAACCTGTTTAAAGATTTATTTAAAGATCTACCTG 540
481 CTTCAATCTTGGTCAAGGTACAACTGTAACCTGTTTAAAGATTTATTTAAAGATCTACCTG 540
541 TAAGAAAGCAGTTTACTCAACTGCAAAAATAAGTAAAGTGAATAAAAAAGATCCAAAG 600
541 TAAGAAAGCAGTTTACTCAACTGCAAAAATAAGTAAAGTGAATAAAAAAGATCCAAAG 600
601 ATCTCTCATGAGCTTTGGTATCCCTTAAACCTGACTTAAAGGATTTGCTTTGTACATAACA 660
601 ATCTCTCATGAGCTTTGGTATCCCTTAAACCTGACTTAAAGGATTTGCTTTGTACATAACA 660
661 AGGCAGTTATTTGGCAGAAAAGCAGATATCAGATCAAGATGGCTCTCATGTGAGTTC 720
661 AGGCAGTTATTTGGCAGAAAAGCAGATATCAGATCAAGATGGCTCTCATGTGAGTTC 720
721 TGGGGACTGCTGTTATGAACATATGGAATCCCTTTCAGTACCACTCTGAAGAAATCTCAGA 780
721 TGGGGACTGCTGTTATGAACATATGGAATCCCTTTCAGTACCACTCTGAAGAAATCTCAGA 780
781 TTTATCTCAGTGAGTTCTTCCAAAGTGTGATGCAGACCACTCTTTCAGTACTTTCCTCAA 840
781 TTTATCTCAGTGAGTTCTTCCAAAGTGTGATGCAGACCACTCTTTCAGTACTTTCCTCAA 840
841 CACCAGAAAGAGTTTCACTTTCATAAAGTGTGATGCAGACCACTCTTTCAGTACTTTCCTCAA 900
841 CACCAGAAAGAGTTTCACTTTCATAAAGTGTGATGCAGACCACTCTTTCAGTACTTTCCTCAA 900
901 AGTTAATCCGACATCATTAACATCTGAAATGCTTAAAGGAATCTACTCGTTTGTATCCTG 960
901 AGTTAATCCGACATCATTAACATCTGAAATGCTTAAAGGAATCTACTCGTTTGTATCCTG 960
961 TTTTCTTCTGAAAATCGATGTTCCCTACAGCTGATGTTGATGTAATTTAAACACAGATA 1020
961 TTTTCTTCTGAAAATCGATGTTCCCTACAGCTGATGTTGATGTAATTTAAACACAGATA 1020
1021 AAAGCCAAAGTATTTAACAATAAGGAATCTGTTTAAATTTGCTTCTTGAATAATCTGATGA 1080
1021 AAAGCCAAAGTATTTAACAATAAGGAATCTGTTTAAATTTGCTTCTTGAATAATCTGATGA 1080
1081 CGACTGTTATGACCAATTAACCTAGTACAAATTTCTTATGAAATTAATAAACAAGATGTTT 1140
1081 CGACTGTTATGACCAATTAACCTAGTACAAATTTCTTATGAAATTAATAAACAAGATGTTT 1140
1141 CCCAGCTGACATGCTTTAGTAAAAACAGCAGAAAACAGATGTGCTTTTAAATAAGTGG 1200
1141 CCCAGCTGACATGCTTTAGTAAAAACAGCAGAAAACAGATGTGCTTTTAAATAAGTGG 1200
1201 AATCATCTGGAAGAATAATTTCAAAATGTTGATATCTTCAGTCAATTCCTCAAAAATGATA 1260
1201 AATCATCTGGAAGAATAATTTCAAAATGTTGATATCTTCAGTCAATTCCTCAAAAATGATA 1260
1261 TGCAATATGATGATCTGGAAAAAACACTGATGATGTTGTTTAAATCACCAGATAGTATTG 1320

1261 TGCATAATGATGATCTGGAAAAAACACTGATGATGTTGTTTAAATCACCAGATAGTATTG 1320
1321 GTGACTTTGGTTATGGTCAATTTGATAGTGAATTTCTTAAATGATGATAAACAACACTAAGA 1380
1321 GTGACTTTGGTTATGGTCAATTTGATAGTGAATTTCTTAAATGATGATAAACAACACTAAGA 1380
1381 ATGCAATTTCAGGACATTTCAATGAGTAATATCATGAGGAGAACTCTCAGACGGAATATA 1440
1381 ATGCAATTTCAGGACATTTCAATGAGTAATATCATGAGGAGAACTCTCAGACGGAATATA 1440
1441 GTAAAACTTGTTTTATAAGTTCGCTTAAAGCAACCCAGCTCAGAAATGCGCAATAAGACC 1500
1441 GTAAAACTTGTTTTATAAGTTCGCTTAAAGCAACCCAGCTCAGAAATGCGCAATAAGACC 1500
1501 ATATAGATGAGAGTGGGAAAAATGAGGAAGAGCAGGCTCTTGAAAACTCTTCGGAATTTT 1560
1501 ATATAGATGAGAGTGGGAAAAATGAGGAAGAGCAGGCTCTTGAAAACTCTTCGGAATTTT 1560
1561 CTGCAGATGAGTGGAGCAGGGGAAATATACTTAAAAATTCAGTGGGAGAGAAATATGAAAC 1620
1561 CTGCAGATGAGTGGAGCAGGGGAAATATACTTAAAAATTCAGTGGGAGAGAAATATGAAAC 1620
1621 CTGTGAAAAATTTTGTGCTGAAAAAAGTTTACCATGTAAGTAATTAATAATATATC 1680
1621 CTGTGAAAAATTTTGTGCTGAAAAAAGTTTACCATGTAAGTAATTAATAATATATC 1680
1681 CAATCCCTGMAAATAATGAAATCTTAAATGAAGATTCATGTAAACAAAAATCAAAATGTAATAG 1740
1681 CAATCCCTGMAAATAATGAAATCTTAAATGAAGATTCATGTAAACAAAAATCAAAATGTAATAG 1740
1741 ATAAATAATCTGAAAAAGTTACAGCTTATGATTTTCTTAGCAATCGAGTAATCAAGAAAC 1800
1741 ATAAATAATCTGAAAAAGTTACAGCTTATGATTTTCTTAGCAATCGAGTAATCAAGAAAC 1800
1801 CCAATGTCAGAAAGTCTCTTTTGTTCAGATCATCGTCTCAGTCTTCTCATAGAAAAATC 1860
1801 CCAATGTCAGAAAGTCTCTTTTGTTCAGATCATCGTCTCAGTCTTCTCATAGAAAAATC 1860
1861 CTAAGACTAGTTTACAGGATGCAACACTACAAATTTGAAGAACTGTGGAAGACATTCAGTGTG 1920
1861 CTAAGACTAGTTTACAGGATGCAACACTACAAATTTGAAGAACTGTGGAAGACATTCAGTGTG 1920
1921 AAGAGAAAAAACTGAAAAATATGAAGAGAGGCTTCTAAAGACTTTGGAACGATACAAATAGTC 1980
1921 AAGAGAAAAAACTGAAAAATATGAAGAGAGGCTTCTAAAGACTTTGGAACGATACAAATAGTC 1980
1981 AAATGAAGAGAGCCATTGAAACAGGAGTCAAAATGTCTCAATAAGATGCGCAAGAAAAAGA 2040
1981 AAATGAAGAGAGCCATTGAAACAGGAGTCAAAATGTCTCAATAAGATGCGCAAGAAAAAGA 2040
2041 TAAAAACCCACAGCGCATGGAATTTGGCCAGAGCAAGTAAATAAACTCTCATTTATCTA 2100
2041 TAAAAACCCACAGCGCATGGAATTTGGCCAGAGCAAGTAAATAAACTCTCATTTATCTA 2100
2101 ATCAACCAAACTTGTGATGAACTCTTTCAGTCCCAATTTGAAAAAGAGAGGTCAAAATA 2160
2101 ATCAACCAAACTTGTGATGAACTCTTTCAGTCCCAATTTGAAAAAGAGAGGTCAAAATA 2160
2161 TTAATAATGATACAGATCCCTCTTCTATGAAAACTTAAATAAATTTTAAAGAAAAA 2220
2161 TTAATAATGATACAGATCCCTCTTCTATGAAAACTTAAATAAATTTTAAAGAAAAA 2220
2221 ACAGAGTTGACCTTAGAAGAGAGGATGAACCTTGTGATCCCAATCTCAGGTTTCTCTG 2280
2221 ACAGAGTTGACCTTAGAAGAGAGGATGAACCTTGTGATCCCAATCTCAGGTTTCTCTG 2280
2281 ATGCAATGCTTAATGAATCCAAAAACAGAGGTAAATGTTATTAATCCATATAGAGTGAAG 2340
2281 ATGCAATGCTTAATGAATCCAAAAACAGAGGTAAATGTTATTAATCCATATAGAGTGAAG 2340
2341 AAGCCCTGCTATTTAAAAAGACTTTCTTGAGAACTTCTTGAGAACTTCTTGAGAACTTCTGAAA 2400
2341 AAGCCCTGCTATTTAAAAAGACTTTCTTGAGAACTTCTTGAGAACTTCTTGAGAACTTCTGAAA 2400